



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

보건학 석사 학위논문

유전자 네트워크 분석을 통한
심근증 마커 유전자 탐색 연구

**RESEARCH ON CARDIOMYOPATHY
MARKER GENE THROUGH GENE
NETWORK ANALYSIS**

2017년 8월

서울대학교 보건대학원
보건학과 생명정보학 전공
서명석

유전자 네트워크 분석을 통한 심근증 마커 유전자 탐색 연구

지도교수 손 현 석

이 논문을 보건학 석사학위논문으로 제출함

2017년 5월

서울대학교 보건대학원

보건학과 생명정보학 전공

서 명 석

서명석의 석사학위논문을 인준함

2017년 6월

위 원 장	<u>조 성 일</u>	(인)
부 위 원 장	<u>정 효 지</u>	(인)
위 원	<u>손 현 석</u>	(인)

국 문 초 록

유전자 네트워크 분석을 통한 심근증 마커 유전자 탐색 연구

서울대학교 보건대학원 보건학과
생명정보학 전공

서 명 석

심근증은 관상동맥질환이나 고혈압, 판막증, 선천성 심장병등의 심혈관계 질환 없이 심장근육에 구조적, 기능적 이상이 생기는 질병으로 심장의 심실이 비정상적으로 비대해 지거나 확장이 되는 특징을 보인다. 심근증은 모든 연령대에서 발병 할 수 있고 증상은 심부전, 부정맥, 혈전증, 심장돌연사 등 다양하게 나타난다. 또한 치료법도 아형, 증상에 따라 매우 다양하기 때문에 효율적인 치료와 관리를 위해서는 조기에 정확하게 진단을 하는 것이 중요하다. 최근 염기서열 분석법 및 생명정보학적 기법의 발달로 심근증과 관련이 있는 유전학적인 정보들이 증가하고 있을 뿐 만 아니라 기존의 염기서열 분석법들에 비해 훨씬 많은 유전자를 빠른 시간에 분석 할 수 있는데, 이는 질병과 관련된 전체적인 유전자 분석 결과를 제공하여 심근증을 진단 하는데 도움을 준다. 이러한 유전학적 정보를 효율적으로 이용하기 위해서는 데이터베이스 및 검색 시스템의 구축이 중요하기 때문에 본 연구에서는 먼저 심근증에 대한 유전학적 데이터베이스와 웹 기반의 검색 시스템을 구축하고 CGDB (Cardiomyopathy Genetic Database)라 명명하였다. 데이터베이스 구축에 이용된 유전자 및 아미노산 서열, 유전자의 변이 및 임상적 의의, 단백질에 대한 정보는 각각 GenBank, ClinVar, UniProt에서 얻었으며 다운로드받은 플랫폼파일은 JAVA 프로그래밍 언어를 이용해 데이터 파싱 과정을

수행하여 데이터베이스의 구축에 필요한 정보들만을 추출하고 이를 다시 Mysql에 입력하기 위한 sql 포맷의 파일로 재가공 하였다. 이렇게 구축한 데이터베이스는 JSP, HTML을 이용하여 웹 기반의 검색 시스템을 통해 구현 하였다. 본 데이터베이스의 구현은 크게 두 가지로 나눌 수 있는데 첫 번째는 심근증에 대한 통합적인 유전학적 정보의 제공으로 조건을 만족하는 유전자를 검색 할 수 있고 해당 유전자의 변이 및 발현하는 단백질에 대한 정보를 얻을 수 있게 하였다. 두 번째는 2차 연구를 위한 염기 및 아미노산의 서열정보를 fasta형식의 파일로 출력 할 수 있게 하였다. 이러한 데이터베이스 및 검색시스템을 통해 제공되는 심근증에 대한 유전학적 정보들은 신속하고 정확한 심근증의 진단에 크게 기여 할 수 있을 것이다. 다음으로는 심근증에 대한 마커 유전자를 찾기 위해 유전자 네트워크 분석을 실시 하였다. 먼저 네트워크를 구성 할 유전자를 선별하기 위해 NCBI GEO의 마이크로어레이 데이터를 이용하였다. 여러 원인의 심근증 환자들 중 심부전의 예후를 나타내는 환자들에 대한 연구 결과들로 차별 및 공통 발현 유전자들을 선별하여 심근증의 예후로 심부전을 나타내는데 관여하는 유전자들의 네트워크를 구성 하였다. 유전자들 사이의 연결선은 유전자들 간의 발현 상관성을 기반으로 구성하였는데 발현 상관성에 대한 정보는 GeneNetwork라는 웹 서비스를 통해 얻었다. 이렇게 구성한 유전자 네트워크에 대해 중심성 분석을 실시하였는데 이를 통해 C1QTNF7, ECM2, FAM188A, HNMT, ITIH5, NAMPT, OIP5-AS1, PTN, SMOC2, SULF1이 심부전의 예후를 나타내는데 중요한 역할을 한다는 것을 알 수 있었다. 또한 위의 네트워크와 선천적인 변이가 심근증의 원인이 되는 유전자들간의 2부 네트워크를 구성하고 중심성 분석을 시행하여 심부전의 예후를 보이게 하는 유전자들과 발현상관성이 큰 선천적 변이의 유전자들을 선별하였다. 중심성 분석의 결과 확장성 심근증 에서는

ACTC1, ACTN2, BAG3, DES, DMD, DSG2, DSP, FKTN, ILK, LAMP2, LDB3, LMNA, MYH7, MYPN, PKD1, PKP2, PLN, PSEN1, RBM20, RYR2, SCN5A, SGCD, TCAP, TPM1, TTN, VCL이, 비대성 심근증 에서는 ACTC1, ACTN2, ANKRD1, LAMP2, MYOZ2, MYPN, PLN, PRKAG2, TTN, VCL이 심부전의 예후를 보이게 하는 유전자들과 발현상관성이 크다는 것을 알 수 있었다. 이는 말초 혈액을 이용한 통상적인 진단검사에 심부전의 예후를 예측 할 수 있는 마커 유전자로 활용이 가능하다. 본 연구 결과들은 심부전에 대한 신속 정확한 진단 및 예후를 예측 할 수 있게 해주어 적절한 치료 방향 결정 및 지속적인 모니터링을 통해 환자들의 수명 연장 및 삶의 질 향상에 기여할 수 있는 보건학적 의의가 있다.

.....
주요어 : 심근증, 생명정보학, 데이터베이스, 네트워크분석, 마커유전자, 보건, 진단검사

학 번 : 2015-24008

부족한 제자를 열정으로 이끌어주신
지도교수님께 이 논문을 바칩니다.

목 차

국문 초록	i
목 차	v
표 목 차	vii
그림 목차	ix
약어 목록	x

제 1 장. 서 론

1.1 연구배경

1.1.1 심근증의 정의와 병리생리학적 특징	1
1.1.2 심근증에 대한 분자 수준의 연구	5
1.1.3 심근증으로 인한 심부전	17
1.1.4 심근증 관련 데이터베이스	18
1.1.5 마이크로어레이와 차별발현 유전자 분석	23
1.1.6 유전자 네트워크 분석	25

1.2 연구의 필요성	27
-------------------	----

1.3 연구의 목적	29
------------------	----

제 2 장. 연구 방법

2.1 데이터 수집 및 가공	31
-----------------------	----

2.2 시스템 개발 환경 및 데이터베이스 구현	35
---------------------------------	----

2.3 차별발현 유전자 분석	39
2.4 유전자 네트워크 분석	43
2.5 단백질 네트워크 분석	54
 제 3 장. 연구 결과	
3.1 심근증에 대한 유전학적 정보 검색 시스템 구축	55
3.2 마이크로어레이 데이터를 이용한 차별발현 유전자 분석 ..	63
3.3 유전자 네트워크 분석	68
3.4 단백질 네트워크 분석	83
3.5 유전자 선별	86
 제 4 장. 고 찰	
4.1 심근증으로 인한 심부전의 마커유전자 탐색	97
4.2 활용 방안	99
4.3 보건학적 연구에 적용	101
 제 5 장. 결 론	102
참고문헌	105
 Abstract	111

표 목 차

Table 1.1 Genes associated with hypertrophic cardiomyopathy	10
Table 1.2 Genes associated with dilated cardiomyopathy	11
Table 1.3 Genes associated with arrhythmogenic right ventricular cardiomyopathy	12
Table 1.4 Genes associated with restrictive cardiomyopathy	13
Table 1.5 Genes associated with left ventricular non-compaction cardiomyopathy	13
Table 1.6 Main genes of interest in cardiomyopathies for mutation screening in routine practice	14
Table 1.7 Statistical data on the frequency of mutations in the genes associated with cardiomyopathy including the number of samples, OR and EF for types of variant	15
Table 1.8 Database deal with information of cardiomyopathy	22
Table 2.1 Table structure of database	38
Table 2.2 GEO Series GSE1145 data that used for this study	41
Table 2.3 Dataset	42
Table 2.4 Sample list of left ventricle in GTEx v5	47
Table 2.5 Genes that have germline variant in cardiomyopathy	53
Table 2.6 List of network for this study	53
Table 3.1 Gene list through the gene searching system	62
Table 3.2 Quantity of DEGs and Co-EGs	65
Table 3.3 Co-regulated genes of comparison4	65
Table 3.4 Major components of Functional annotation	66
Table 3.5 Centrality analysis of network2	74
Table 3.6 Important genes of network2	76

Table 3.7 Two-mode network centrality analysis of network3 about germline variant of dilated cardiomyopathy (high ranked 30 genes)	77
Table 3.8 Two-mode network centrality analysis of network3 about comparison4 genes	78
Table 3.9 Important genes of network3	79
Table 3.10 Two-mode network centrality analysis of network4 about germline variant of hypertrophic cardiomyopathy	80
Table 3.11 Two-mode network centrality analysis of network4 about comparison4 genes	81
Table 3.12 Important genes of network4	82
Table 3.13 Two-mode network centrality analysis of network5 about germline variant of dilated cardiomyopathy	84
Table 3.14 Two-mode network centrality analysis of network5 about germline variant of hypertrophic cardiomyopathy	85
Table 3.15 The sum of each centrality indicator and OMIM database search results	88
Table 3.16 Genes that can predict heart failure with DCM prognosis	93
Table 3.17 Genes that can predict heart failure with HCM prognosis	95
Table 3.18 Genes that can predict heart failure with low adjusted centrality.....	96

그림 목 차

Figure 2.1 Genbank flat file and sql file format	33
Figure 2.2 ClinVar flat file and sql file format	33
Figure 2.3 UniProt flat file and sql file format	34
Figure 2.4 Fasta format file which download from database	37
Figure 3.1 Front page of database	57
Figure 3.2 Medical information page	57
Figure 3.3 List for subtype of cardiomyopathy	58
Figure 3.4 List for gene variant	58
Figure 3.5 Information about the protein structure	59
Figure 3.6 Search page	60
Figure 3.7 Genetic information that satisfying options of Search page	60
Figure 3.8 Gene list for download about sequence data	61
Figure 3.9 Sequence data about selected gene	61
Figure 3.10 Non-weighted and directional network that composed of comparison4 genes	72
Figure 3.11 Two-mode network between Comparison4 genes and germline variant of dilated cardiomyopathy	73
Figure 3.12 Two-mode network between Comparison4 genes and germline variant of hypertrophic cardiomyopathy	73
Figure 3.13 Logistic regression curve for gene selection	91
Figure 3.14 ROC curve for logistic regression curve	92

약 어

ACTN2	actinin alpha 2
ANKRD1	ankyrin repeat domain 1
ARVC	arrhythmogenic right ventricular cardiomyopathy
AUC	area under the curve
BAG3	BCL2 associated athanogene 3
BLAST	basic local alignment search tool
CDS	coding sequence
CGDB	Cardiomyopathy Genetic Database
CHD	coronary heart disease
DAVID	the Database for Annotation, Visualization and Integrated Discovery
DCM	dilated cardiomyopathy
DDBJ	DNA Data Bank of Japan
DEGs	differentially expressed genes
DES	desmin
DMD	dystrophin
DSG2	desmoglein 2
DSP	desmoplakin
EBI	the European Bioinformatics Institute
EKG	electrocardiogram
EMBL	European Molecular Biology Laboratory
FDR	false discovery rate
FKTN	fukutin
FWER	family wise error rate
GEO	gene expression omnibus
GI	GenInfo Identifier
GN	GeneNetwork
GO	Gene ontology
GTE_x	the Genotype-Tissue Expression
HCM	hypertrophic cardiomyopathy
HTML	hypertext markup language
ICM	ischemic cardiomyopathy

ILK	integrin linked kinase
JHUSOM	Johns Hopkins University School of Medicine
JSP	Java server page
LAMP2	lysosomal associated membrane protein 2
LDB3	LIM domain binding 3
LMNA	lamin A/C
LVNC	left ventricular non-compaction cardiomyopathy
MAS5	Microarray Suite 5.0
MIM	Mendelian Inheritance in Man
MYBPC3	myosin binding protein C
MYH7	myosin heavy chain 7
MYL3	myosin light chain 3
MYOZ2	myozenin 2
MYPN	myopalladin
MySQL	My sequel
NCBI	National Center for Biotechnology Information
NIH	National Institutes of Health
NGS	next generation sequencing
OMIM	Online Mendelian Inheritance in Man
PIR	the Protein Information Resource
PKP2	plakophilin 2
PLN	phospholamban
PSEN1	presenilin 1
RBM20	RNA binding motif protein 20
RCM	restrictive cardiomyopathy
ROC	receiver operating characteristic
RYR2	ryanodine receptor 2
SCN5A	sodium voltage-gated channel alpha subunit 5
SGCD	sarcoglycan delta
SIB	the Swiss Institute of Bioinformatics
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TCAP	titin-cap
TPM1	tropomyosin 1
TTN	titin
VCL	vinculin

제 1 장. 서 론

1.1 연구 배경

1.1.1 심근증의 정의와 병리생리학적 특징

심장은 올바른 기능을 하기 위해서 심장을 이루고 있는 수축성의 심장근육세포가 내피세포(endothelial cells), 혈관평활근(vascular smooth muscle), 섬유아세포(fibroblast) 그리고 면역세포와 네트 워크를 형성하고 있으며 심장근육의 수축은 간극연접(gap junction)을 통해 전기 화학적으로 조절이 된다. 이러한 수축의 기본단위는 근절(sarcomere)로 이는 20여개의 단백질로 구성되어 있으며 그 외에 존재하는 다양한 단백질들이 심장근육세포의 수축에 관여를 하는데 이 단백질들 간의 상호작용에 이상이 생기면 심장에 질병이 생기게 된다(Harvey and Leinwand, 2011).

심근증(Cardiomyopathy)은 관상동맥질환(coronary artery disease)이나 고혈압(hypertension), 판막증(valvular heart disease), 선천성 심장병(congenital heart disease)등의 심혈관계 질환 없이 심장근육에 구조적, 기능적 이상이 생기는 이질성(heterogeneous) 질병군으로, 심장의 심실이 비정상적으로 비대해 지거나 확장이 되는 특징을 보인다(Elliott et al., 2007). 심근증의 증상으로는 호흡장애(shortness of breath or trouble breathing), 피로(fatigue), 어지럼증(dizziness), 신체활동 중 기절(fainting), 흉통(chest pain)등이 있으며 때에 따라서는 증상이 나타나지 않는 경우도 있다. 심근증은 심부전(heart failure)을 동반하기도 하는데 이는 심장기능의 저하로 신체에 혈액을 정상적으로 공급 해

주지 못하는 데서 비롯 된다(McNally et al., 2015). 또한 심근증은 심장돌연사(sudden cardiac death)와 심부정맥(cardiac arrhythmia)의 주요한 원인이다(McKenna et al., 1981; Maron et al., 2003). 특히 아동기 심장돌연사의 중요한 원인 중 하나가 심근증인데 이는 태어날 당시에는 아무런 증상을 보이지 않는 경우가 많고 그 후에는 매우 다양한 증상을 나타내기 때문에 아직까지 심근증으로 인한 아동기 심장돌연사의 명확한 위험 인자가 밝혀지지 않았다(Bharucha et al., 2015; Deshpande et al., 2016).

심근증은 심근의 구조적, 기능적 변화에 따라 비대성 심근증(hypertrophic cardiomyopathy, HCM), 확장성 심근증(dilated cardiomyopathy, DCM), 제한성 심근증(restrictive cardiomyopathy, RCM), 부정맥 유발성 우심실 심근증(arrhythmogenic right ventricular cardiomyopathy, ARVC), 좌심실 비치밀 심근증(Left ventricular non-compaction cardiomyopathy, LVNC), 그리고 위의 분류에 속하지 않는 그 외의 심근증(unclassified cardiomyopathys)으로 분류 할 수 있다(Charron et al., 2010; Towbin et al., 2015). 비대성 심근증은 근절 유전자의 변이가 대표적인 원인인 단일 유전자성(monogenic) 심장질환이다. 좌심실의 비대(left ventricular hypertrophy), 근섬유혼란(myofibrillar disarray), 심근섬유증(myocardial fibrosis)이 특징적으로 나타나는데, 대부분의 환자에서 증상이 나타나지는 않지만 개인에 따라서 심부전이 빠르게 진행되거나 심장돌연사가 나타나는 등 좋지 않은 예후를 보이기도 한다(Kramer et al., 2015). 확장성 심근증은 좌심실의 수축성 기능장애와 좌심실의 확장(left ventricular dilatation)이 특징적으로 나타나는 유전성 질환으로 서양환자들 중 최소 25% 정도가 상염색체 우성으로 유전되는 형태를 보인다(Elliott et al., 2007). 또한 연간 영, 유아 100,000명당 1.13명의 비율로 발생하며, 소아에서의 심장이식과 심부전의 주된 원인이 되는 등 소아에서 나타나는 심근증 중

가장 많은 비율을 차지 한다(Miranda et al., 2015). 제한성 심근증은 심실이 비후해지거나 확장되지 않는 등 심실의 크기는 정상이지만 혈류동력학적 기능이 손상되어 심장 내 혈압이 증가하는데 이로 인해 심장의 수축기에는 이상이 없지만 이완기에는 기능장애가 나타난다. 호흡장애와 심계항진 (palpitation)이 흔하게 나타나는 증상이며 발병 원인에 따라 다양한 예후가 나타나는데, 발병 원인으로는 특발성, 유전성, 아밀로이드증 이나 유육종증, 피부경화증 같은 전신성 질환 등으로 다양하다(McCartan et al., 2012; Elliott et al., 2007). 부정맥 유발성 우심실 심근증은 심장의 구조적 변화로 인해 부정맥과 비정상적인 심벽운동 등을 일으키는데 이는 우심실의 심근이 섬유성 조직으로 대체되어 일어난다. 주로 유전자 변이에 의해 이 질병이 유발되는데 칼슘통로, 데스모솜 단백질, 지방 생성 경로에 관련된 유전자가 여기에 속한다(Deshpande et al., 2016). 부정맥 유발성 우심실 심근증은 운동선수들이나 유아들에서 나타나는 심장돌연사의 주요한 원인인데, 증상 없이 발병하는 경우가 많아 사후 부검에서 이 질병을 진단받는 경우가 있다(Saguner et al., 2015). 좌심실 비치밀 심근증은 좌심실에 비정상적인 세포주(trabeculae)가 생겨서 심장의 수축성(systolic) 혹은 확장성(diastolic) 기능장애가 발생하는 질병이다. 선천적으로 심장의 발생 과정 중 이상으로 질병이 발생 할 것이라 생각 하지만 아직 확실한 발병 기전은 밝혀지지 않았다. 어느 연령대에서나 발병 할 수 있고, 아무런 증상이 없는 것부터 심부전, 부정맥, 심장돌연사, 혈전색전증에 이르기까지 다양한 증상이 나타난다(Towbin et al., 2015). 이 밖에도 위의 기준으로 분류되지 않는 스트레스성 심근증(stress cardiomyopathy), 허혈성 심근증(Ischemic cardiomyopathy, ICM)등이 있다(Elliott et al., 2007).

심근증의 진단은 개인과 가족의 의무기록, 신체검사, 진단검사 등을 통하여 이루어 진다. 제일 처음 의무기록과 현재 증상과 징후

를 파악 한 뒤 심부전이나 심장돌연사에 대한 가족력을 확인한다. 다음으로는 청진기로 심장과 폐의 소리는 듣는다. 그리고 다양한 진단 검사들을 하는데 혈액검사를 통해서 심장에 대한 정보를 얻는 것과 동시에 심근증 이외의 다른 질병을 제외 시킬 수 있고, 흉부 엑스선 검사(chest X ray)를 통해서 심장의 확장 유무를 알 수 있다. 심전도검사(electrocardiogram, EKG)를 통해서 심장의 전기적 활성을 알 수 있는데 심박수와 리듬의 이상 유무를 확인 할 수 있다. 또한 심장 각 부위의 전기신호 전달 속도도 알 수 있다. 이러한 심전도 검사는 심장마비, 부정맥, 심부전 등의 다른 심장 질환을 발견하는데도 효과적이다. 심장초음파(echocardiography)는 음파를 이용해서 심장이 움직이는 영상을 얻을 수 있는데 이를 통해서 심장의 크기, 형태, 운동이 정상적인지를 알 수 있다. 유전자 검사는 심근증에 대한 가족력이 있는 경우에 유용하게 사용 될 수 있는데, 이는 질병의 증상이나 징후가 나타나기 전에 심근증의 유무를 판단 할 수 있게 해 주기 때문에 조기에 치료를 할 수 있게 해 준다(Elliott et al. 2014).

1.1.2 심근증에 대한 분자 수준의 연구

최근 심근증과 관련이 있는 많은 유전자들이 발견되고 있는데, 각각의 유전자가 가지는 다양한 변이의 결과로 다양한 증상이 나타난다. 뿐만 아니라 같은 변이에도 경우에 따라 증상의 정도가 다르게 나타나기도 하는데, 예를 들어 부정맥 유발성 우심실 심근증 환자의 12%, 확장성 심근증 환자의 15%에서 관찰 되는 포스포람반(phospholamban) 유전자의 변이는 어린 나이에 치명적인 부정맥을 동반하는 심근증을 일으키기도 하는데, 이 유전자의 변이를 가지고 있는 확장성 심근증 환자 중에는 늦은 나이가 되서야 증상이 나타나는 경우도 있다(van Rijsingen et al., 2014; DeWitt et al., 2006). 심근증의 발병과 관련이 있는 여러 유전자 변이의 종류와 특성은 심근의 구조적, 기능적 변화를 기준으로 한 심근증의 아형에 따라 다음과 같이 살펴 볼 수 있다.

유전성 비대성 심근증 환자의 약80%는 myosin heavy chain 7 (MYH7) 유전자와 myosin binding protein C (MYBPC3) 유전자의 상염색체 우성 돌연변이를 가지고 있다(Kensler et al., 2011). 이 두 유전자는 각각 근절의 굵은 필라멘트(sarcomere thick filament) 단백질과 심장 미오신 결합 단백질 C에 대한 유전정보를 가지고 있는데, 두 유전자 모두 비대성 심근증과 높은 연관성이 있지만 발병에 작용하는 기전은 서로 다르다. MYH7의 변이는 액틴 결합(actin binding)도메인, 아데노신트리포스파타아제(ATPase)도메인, 힘을 전달하는(force transmission)도메인 등 중요 부위의 아미노산을 다른 아미노산으로 치환시켜 심장의 기능이 저하되게 한다(Walsh et al., 2010). 반면 MYBPC3의 변이는 주로 종결 코돈이 조기에 발현(nonsense mutation)되거나 틀 이동(frame shifting)을 일으켜 온전한 단백질을 형성하지 못하게 하여 단백질이 정상적으로 기능을 못하게 한다. 두 유전자의

변이를 비교해 보면 MYBPC3의 변이에 의한 심근증은 MYH7의 변이에 의한 것 보다 질병이 늦은 나이에 유발되고 증상이 가볍다 (Charron et al., 1998; Maron et al., 2001). 그 외에 비대성 심근증과 관련이 있는 유전자와 이 유전자가 발현하는 단백질, 단백질의 기능은 <Table 1.1>에서 확인 할 수 있다.

유전성 확장성 심근증에서는 Titin (TTN) 유전자의 변이가 많이 나타나는데, 약 25%의 유전성 확장성 심근증 환자가 이 변이를 가지고 있다. Titin 단백질은 가로무늬 단백질 중 세 번째로 풍부하며 약 33,000여개의 아미노산으로 이루어져 있다 (Trinick et al., 1984). 또한 근섬유 분절 조립(sarcomere assembly)에 필수적 이고 수축력을 조절하는 기능이 있는데, 이러한 titin의 변이는 심근과 골격근을 포함한 선천적인 근육병, 호흡 부전(respiratory failure)을 동반하는 유전성 근육병을 야기한다 (Lange et al., 2005). 주로 넉센스 변이, 틀 이동 변이, 스플라이싱 변이, 복제수 변이(copy number variation)가 일어나며, 이로 인해 직렬삽입(tandem insertion)이나 단백질 절삭(protein truncation)이 일어나 titin의 구조가 변하게 된다 (Herman et al., 2012). 그 외에 확장성 심근증과 관련이 있는 유전자와 이 유전자가 발현하는 단백질, 단백질의 기능은 <Table 1.2>에서 확인 할 수 있다.

부정맥 유발성 우심실 심근증은 주로 심근세포간 접합부인 개재판(intercalated disc)의 데스모솜 단백질의 기능 이상으로 일어나는데, 부정맥 유발성 우심실 심근증 환자의 약 30~50%에서 데스모솜 단백질의 변이가 관찰된다. plakophilin-2 (PKP2), desmoglein-2 (DSG2), desmoscollin-2 (DSC2), desmoplakin (DSP), junctional plakoglobin (JUP) 유전자들이 데스모솜 단백질의 변이와 관련이 있는 유전자들인데, 이중 PKP2 유전자의 변이에 의한 경우가 제일 많다 (Marcus et al., 2013). Plakophilin-2는 plakoglobin, desmoplakin과 함께 데스모솜을 이루는데, 이 단백질 복합체는 카데린(cadherin)과 연결되어 인접세포들

과 구조적, 기능적 안정성을 유지시켜 준다. 이러한 구조는 세포간 신호전달에 매우 중요한데, PKP2유전자의 변이는 이러한 세포간의 상호작용에 이상이 생기게 하여 심근증을 유발한다(van Tintelen et al., 2006). 그 외에 부정맥 유발성 우심실 심근증과 관련이 있는 유전자와 이 유전자가 발현하는 단백질, 단백질의 기능은 <Table 1.3>에서 확인 할 수 있다.

제한성 심근증에서는 Troponin I Type 3 (TNNI3) 유전자의 변이가 가장 많이 관찰 되는데 TNNI3 유전자는 얇은 섬유 단백질의 하나인 cardiac troponin I (cTnI)를 암호화 하고 있다(Kapoor et al., 2015). 이 cTnI는 troponin C, troponin T와 함께 트로포닌 복합체를 형성 하는데, 트로포닌 복합체는 심장 근육의 섬유간 상호작용을 조절하여 심장 근육의 수축과 이완을 조절하고 이중 cTnI는 트로포닌 복합체의 입체 구조를 변화시켜 근육의 수축을 유도한다. 트로포닌 복합체의 변이는 칼슘이온의 친화도와 단백질간 상호작용에 영향을 미쳐 심근증을 유발하는데, 이중 cTnI의 변이는 근절의 신축성을 변화 시킨다(Tardiff, 2011). 그 외에 제한성 심근증과 관련이 있는 유전자와 이 유전자가 발현하는 단백질, 단백질의 기능은 <Table 1.4>에서 확인 할 수 있다.

좌심실 비치밀 심근증은 주로 X염색체 열성(X-linked recessive) 혹은 상염색체 우성(autosomal dominant)으로 유전된다. 선천적 심장 질환과 동반되는 좌심실 비치밀 심근증에서는 DTNA와 MYH7 유전자에서 변이를 보이고, 선천적 심장질환을 동반하지 않는 좌심실 비치밀 심근증에서는 Z선(Z-line)을 암호화하고 있는 LDB3 유전자와 TPM1 유전자 등에서 변이를 보인다. 또한 세포골격단백질인 dystrophin을 암호화하고 있는 DMD유전자와, 나트륨통로(sodium channel) 유전자인 SCN5A도 좌심실 비치밀 심근증과 관련이 있다고 알려져 있다(Towbin et al., 2015). 그 외에 좌심실 비치밀 심근증과 관

련이 있는 유전자와 이 유전자가 발현하는 단백질, 단백질의 기능은 <Table 1.5>에서 확인 할 수 있다.

이렇게 심근증과 연관이 있는 유전 변이들에 대한 연구 결과가 증가함에 따라 질병의 진행이나 합병증에 대한 더 많은 정보를 얻을 수 있을 것이다. 예를 들면 LMNA 유전자의 변이는 확장성 심근증에서 두 번째로 흔한 유전 변이인데, 이러한 변이가 있으면 합병증으로 부정맥이 오기 쉽다. 또한 LMNA 유전자에 틀 이동 변이와 좌심실 기능 저하가 있으면 심실 빈맥을 나타내는 경향이 있는데 이러한 정보들은 치료 방법을 계획하는데 도움을 줄 수 있다. 그 외에 심근증의 유전학적 분류는 치료에도 좋은 결과를 가져 올 수 있다. 고속대량스크린(high throughput screen)을 통해 치료의 타겟이 되는 단백질을 찾을 수 있고, 역배열 올리고뉴클레오티드(anti-sense oligonucleotides)를 이용한 유전자 치료에도 위의 연구 결과가 이용될 수 있다(McNally and Puckelwartz, 2015).

이렇게 축적된 심근증과 관련된 유전정보와 차세대 염기서열 분석법(next generation sequencing, NGS)등 발전된 염기서열분석법을 통해 단시간에 대량의 염기서열을 분석 할 수 있게 됨으로써 여러 기관에서 심근증에 대한 전체적인 유전자 검사를 시행 할 수 있게 되었다(Limongelli et al., 2012). <Table 1.6>에는 유전 변이가 원인인 심근증을 선별검사 하기 위한 주요 유전자를 나타내었다. 이러한 심근증에 대한 유전자 검사는 최근까지 주로 연구를 위한 목적으로 시행 되었지만, 심근증 환자 개인에 대해 더 특이적인 정보를 제공 함으로서 진단과 예후 평가, 치료 수단을 결정하는데 도움을 줄 수 있을 뿐만 아니라 환자의 가족들에 대해 추가적으로 심근증의 발병 가능성을 평가 할 수 있기 때문에 점차 임상적으로 이용 하려는 움직임이 있다 (Charron P et al., 2010).

염기서열 분석법의 발전으로 여러 유전자들의 변이 여부를 단

시간에 분석 할 수 있게 됨에 따라 기존의 연구를 통해서도 알기 힘들었던 여러 통계적 분석 결과 등을 얻을 수 있게 되었다. 최근 한 연구에서는 심근증 환자 7,855명의 염기 서열과 Exome Aggregation Consortium (ExAC)의 reference 서열을 비교하여 심근증 환자들에게서 변이가 나타나는 유전자들 각각의 변이 빈도 및 변이 수준에서의 오즈비 등 여러 통계적 결과들을 제시 해 주고 있다 <Table 1.7>. 이러한 결과는 기존에 단순히 유전성 질병의 원인이 되는 유전자들을 선별하는 수준이 아니라 어떠한 유전자가 유전성 질병을 일으키는데 더 큰 영향을 미치는지 알 수 있게 해 준다. 뿐만 아니라 유전자 변이의 종류에 따른 여러 통계적 분석 결과들을 제시해 줌으로서 변이의 종류에 따른 질병과의 연관성도 예측 할 수 있게 해 준다(Walsh et al., 2016). 이는 유전자 검사에 대한 정확한 해석을 가능하게 해주어 유전학적 정보들을 진단검사에 이용하는데 도움을 줄 수 있다.

Table 1.1 Genes associated with hypertrophic cardiomyopathy

Gene	Protein	Function
β -MHC	β -Myosin heavy chain	Sarcomere protein
α -MHC	α -Myosin heavy chain	Sarcomere protein
cMYBPC	Cardiacmyosin-binding protein C	Sarcomere protein
cTnI	Cardiac troponin I	Sarcomere protein
cTnT	Cardiac troponin T	Sarcomere protein
cTnC	Cardiac troponin C	Sarcomere protein
α -TM	α -Tropomyosin	Sarcomere protein
MLC-1	Myosin essential light chain	Sarcomere protein
MLC-2	Myosin regulatory light chain	Sarcomere protein
ACTC	Actin	Sarcomere protein
TTN	Titin	Sarcomere protein
PRKAG2	AMP kinase	
LAMP2	Lysosomemembrane protein	

There are over 500 mutations in 13 genes that have been identified that cause HCM and 50% of these are familial (McCartan et al., 2012).

Table 1.2 Genes associated with dilated cardiomyopathy

Gene	Protein	Function
ACTC	Cardiac actin	Sarcomere protein
DES	Desmin	Dystrophin-associated glycoprotein complex
SGCD	δ -Sarcoglycan	Dystrophin-associated glycoprotein complex
MYH7	β -Myosin heavy chain	Sarcomere protein
TNNT2	Cardiac troponin T	Sarcomere protein
TPM1	α -Tropomyosin	Sarcomere protein
TTN	Titin	Sarcomere structure
VCL	Metavinculin	Intercalated discs
MYBPC	Myosin-binding protein C	Sarcomere protein
MLP/CSRP3	Muscle LIM protein	Z discs
ACTN2	α -Actinin-2	Sarcomere structure
MYH6	α -Myosin heavy chain	Sarcomere protein
ABCC	SUR2A	Cardiac K channel
LMNA	Lamin A/C	Nuclear membrane protein
PLN	Phospholamban	Sarcoplasmic reticulum Ca regulator
ZASP/LBD3	Cypher	Cytoskeletal assembly
DMD	Dystrophin	Dystrophin-associated glycoprotein complex
TAZ/G4.5	Tafazzin	Dystrophin-associated glycoprotein complex
TNNI3	Troponin I	Sarcomere protein

Around 30– 50% of cases have a familial component, and more than 30 genes have been identified, to date, that cause DCM (McCartan et al., 2012).

Table 1.3 Genes associated with arrhythmogenic right ventricular cardiomyopathy

Gene	Protein	Function
TGFB3	Transforming growth factor β 3	Cell signalling
RYR2	Ryanodine receptor 2	Sarcoplasmic reticulum calcium channel
LAMR1	Extracellularmatrix glycoprotein	Cell signalling, adhesion, and migration
PTPLA	Protein-tyrosine phosphatase-like member A	Fatty acid synthesis
DES; ZASP	Desmosomal protein; PDZ domain protein	Dystrophin-associated glycoprotein complex, and cytoskeletal assembly
DSP	Desmoplakin	Anchoring of intermediate filament
PKP2	Plakophilin 2	Cell adhesion
DSG2	Desmoglein 2	Calcium-binding transmembrane glycoprotein
DSC2	Desmocollin 2	Calcium-dependent glycoprotein

It is a familial disease in around 50% of cases and is usually transmitted in an autosomal dominant fashion (McCartan et al., 2012).

Table 1.4 Genes associated with restrictive cardiomyopathy

Gene	Protein	Function
MYL2	Myosin, Light Chain 2	Contractile protein
MYL3	Myosin, Light Chain 3	Regulatory light chain of myosin
MYPN	Myopalladin	Sarcomere protein
ACTC1	Actin Cardiac Muscle 1	Cell motility
MYH7	Myosin, Heavy Chain 7	Sarcomere protein
TNNT2	Troponin T Type 2	Sarcomere protein
TPM1	Tropomyosin 1	Sarcomere protein
DES	Desmin	Cytoskeletal assembly
TNNI3	Troponin I Type 3	Sarcomere protein

Familial restrictive cardiomyopathies are usually inherited in an autosomal dominant fashion (Van Tintelen et al., 2014).

Table 1.5 Genes associated with left ventricular non-compaction cardiomyopathy

Gene	Protein	Function
DTNA	Dystrobrevin alpha	Clustering of nicotinic acetylcholine receptors
MYH7	Myosin, Heavy Chain 7	Sarcomere protein
LDB3	LIM domain-binding protein 3	Adapter in striated muscle
TPM1	Tropomyosin 1	Sarcomere protein
DMD	Dystrophin	Dystrophin-associated glycoprotein complex
SCN5A	Sodium channel protein type 5 subunit alpha	Mediates the voltage-dependent sodium ion permeability

Genetic inheritance arises in at least 30-50% of patients, and several genes that cause left ventricular non-compaction have been identified. These genes seem generally to encode sarcomeric (contractile apparatus) or cytoskeletal proteins (Towbin et al., 2015).

Table 1.6 Main genes of interest in cardiomyopathies for mutation screening in routine practice

Cardiomyopathy	Genes
HCM	MYBPC3 (myosin-binding protein C) MYH7 (beta myosin heavy chain), TNNT2 (troponin T) TNNI3 (troponin I) MYL2 (regulatory myosin light chain) TPM1 (alpha tropomyosin) MYL3 (essential myosin light chain) ACTC (actin) PRKAG2 (AMP-activated protein kinase) LAMP2 (lysosome-associated membrane protein2) GLA (alpha galactosidase), mitochondrial DNA
DCM	MYH7 (beta myosin heavy chain) TNNT2 (troponin T), LMNA (lamin A/C) TAZ (tafazzin), DES (desmin), DMD (dystrophin) mitochondrial DNA
RCM	TNNI3 (troponin I), DES (desmin) TTR (transthyretin), HFE (haemochromatosis)
ARVC	PKP2 (plakophilin2), DSP (desmoplakin) DSG2 (desmoglein2), DSC2 (desmocollin2)
LVNC	MYH7 (beta myosin heavy chain)

HCM : hypertrophic cardiomyopathy, DCM : dilated cardiomyopathy,

RCM : restrictive cardiomyopathy, ARVC : arrhythmogenic right ventricular cardiomyopathy,

LVNC : left ventricular non-compaction cardiomyopathy

At present, the most frequent genes of interest are screened for a mutation using various methods (Charron P, 2010).

Table 1.7 Statistical data on the frequency of mutations in the genes associated with cardiomyopathy including the number of samples, OR and EF for types of variant

Gene	Cases	Case freq.	Non-truncating variants			Truncating variants		
			Case excess	OR	EF	Case excess	OR	EF
< Dilated cardiomyopathy >								
ABCC9	590	0.01356	0.00386	1.49	0.33	0.00095	2.33	0.57
ACTC1	1103	0.00453	0.00301	5.92	0.83	0.00087	27.31	0.96
ACTN2	895	0.00894	-0.00182	0.83	-	-0.00010	5.19	-
ANKRD1	426	0.00000	-0.00304	0.38	-	-0.00052	2.24	-
CASQ2	121	0.01653	0.00364	1.80	0.44	0.00780	18.50	0.95
CAV3	121	0.00000	-0.00124	3.28	-	-0.00014	29.17	-
CRYAB	425	0.00471	0.00273	2.38	0.58	-0.00010	10.79	-
CSRP3	945	0.00423	0.00029	1.10	0.09	0.00070	2.92	0.66
CTF1	590	0.00169	0.00129	4.27	0.77	-0.00006	11.37	-
DES	894	0.01007	0.00347	1.79	0.44	0.00188	7.16	0.86
DSC2	427	0.00468	-0.00470	0.50	-	-0.00026	4.26	-
DSG2	427	0.00703	-0.00523	0.57	-	-0.00072	1.60	-
DSP	427	0.07494	0.01606	1.55	0.35	0.02740	41.01	0.98
DTNA	121	0.00826	0.00400	1.94	0.49	-0.00010	38.12	-
EMD	590	0.00169	-0.00061	1.02	-	0.00000	-	-
FHL1	355	0.00000	-0.00170	1.14	-	0.00000	-	-
FHL2	425	0.00235	-0.00065	0.78	-	-0.00010	10.71	-
GLA	541	0.00000	-0.00138	0.92	-	0.00000	-	-
JUP	425	0.00471	-0.00709	0.40	-	-0.00018	6.25	-
LAMA4	121	0.04959	0.02779	2.34	0.57	-0.00102	4.06	-
LAMP2	532	0.00376	0.00102	1.90	0.47	0.00000	-	-
LDB3	740	0.01486	0.00412	1.39	0.28	-0.00048	1.33	-
LMNA	1044	0.04406	0.02264	4.83	0.79	0.01519	99.72	0.99
MYBPC3	1161	0.02498	0.00614	1.33	0.25	-0.00086	0.50	-
MYH6	121	0.02479	0.00143	1.06	0.06	-0.00128	3.24	-
MYH7	1315	0.05247	0.03821	3.99	0.75	0.00028	1.59	0.37
MYL2	543	0.00184	0.00030	1.20	0.17	-0.00024	3.84	-
MYL3	543	0.00000	-0.00178	0.51	-	-0.00004	22.30	-
MYLK2	121	0.01653	0.00885	2.17	0.54	-0.00014	27.70	-
MYOZ2	121	0.00000	-0.00236	1.73	-	-0.00020	19.89	-
NEXN	156	0.03205	0.01872	3.78	0.74	0.00577	10.06	0.90
PKP2	427	0.01639	0.00359	1.28	0.22	-0.00076	1.57	-
PLN	1095	0.00365	0.00232	6.64	0.85	0.00085	13.82	0.93
PRKAG2	546	0.00366	-0.00154	0.70	-	-0.00012	7.04	-
RBM20	156	0.05769	0.04375	4.33	0.77	-0.00018	12.84	-
RYR2	121	0.04959	0.01589	1.49	0.33	-0.00114	3.71	-
SCN5A	304	0.02303	-0.01002	0.56	-	0.00923	16.53	0.94

Table 1.7 Statistical data on the frequency of mutations in the genes associated with cardiomyopathy including the number of samples, OR and EF for types of variant (cont.)

			Non-truncating variants			Truncating variants		
Gene	Cases	Case freq.	Case excess	OR	EF	Case excess	OR	EF
< Dilated cardiomyopathy >								
SGCD	590	0.00678	0.00194	1.61	0.38	0.00147	8.61	0.88
TAZ	740	0.00405	0.00144	2.96	0.66	0.00133	69.45	0.99
TCAP	590	0.01186	0.00968	5.51	0.82	-0.00024	3.22	-
TMEM43	427	0.00703	0.00013	1.02	0.02	-0.00040	2.87	-
TNNC1	156	0.01923	0.01869	35.13	0.97	-0.00004	75.60	-
TNNI3	1239	0.00646	0.00460	3.46	0.71	-0.00042	0.97	-
TNNT2	1254	0.02871	0.02657	13.88	0.93	-0.00030	1.30	-
TPM1	1111	0.01890	0.01626	20.84	0.95	0.00176	52.95	0.98
TTN	460	0.14565	-	-	-	0.13689	19.77	0.95
TTR	121	0.00000	-0.00136	3.02	-	-0.00004	71.29	-
VCL	590	0.02203	0.00541	1.56	0.36	0.00646	21.33	0.95
< Hypertrophic cardiomyopathy >								
ACTC1	4185	0.00526	0.00464	8.59	0.88	-0.00004	2.88	-
ACTN2	1439	0.00695	-0.00451	0.58	-	0.00059	7.00	0.86
ANKRD1	807	0.00496	0.00068	1.23	0.19	0.00072	2.41	0.59
CSRP3	2167	0.00508	0.00081	1.28	0.22	0.00102	3.82	0.74
FHL1	1535	0.00912	0.00724	6.95	0.86	0.00065		-
GLA	3700	0.00730	0.00630	7.39	0.86	0.00000		-
LAMP2	3290	0.00638	0.00045	1.23	0.19	0.00395		-
MYBPC3	6179	0.19032	0.07988	5.70	0.82	0.09074	118.7	0.99
MYH7	6112	0.14218	0.12786	12.04	0.92	0.00034	1.71	0.42
MYL2	4185	0.01099	0.00873	6.74	0.85	0.00048	3.10	0.68
MYL3	4185	0.00884	0.00706	5.00	0.80	-0.00004	2.90	-
MYOZ2	632	0.00158	-0.00078	0.67	-	-0.00020	3.82	-
NEXN	632	0.01266	0.00416	1.61	0.38	0.00094	2.47	0.60
PLN	2167	0.00323	0.00096	3.35	0.70	0.00179	27.96	0.96
PRKAG2	3973	0.01057	0.00487	1.95	0.49	0.00038	4.15	0.76
TNNC1	632	0.00000	-0.00054	1.40	-	-0.00004	18.70	-
TNNI3	6047	0.02233	0.01947	11.60	0.91	0.00057	2.43	0.59
TNNT2	6103	0.01950	0.01556	8.46	0.88	0.00150	6.04	0.83
TPM1	4447	0.01484	0.01400	18.04	0.94	-0.00004	2.64	-
TTR	632	0.00316	0.00180	2.34	0.57	-0.00004	13.69	-

Case frequency = Cases with gene variant) / (Cases with gene variant + Cases with gene variant)

Case excess = Case frequency - ExAC frequency

OR : odds ratio, EF : etiological fraction = (OR-1)/OR × 100

Non-truncating variants : missense, in-frame insertions and deletions

Truncating variants : frameshift, nonsense, splice donor/acceptor

(Walsh et al., 2016)

1.1.3 심근증으로 인한 심부전

심근증의 주요한 예후 중 하나인 심부전은 심장의 구조적 또는 기능적 상으로 인해 심장이 혈액을 받아드리는 이완기능이나 혈액을 내보내는 수축기능이 감소하여 신체 조직에 필요한 혈액을 제대로 공급하지 못해서 발생하는 질환군 이다(Go et al., 2014). 심부전은 관상동맥질환, 고혈압 및 심근증에 의해서 발병하는데, 특히 심근증과 연관이 있는 유전자 변이는 심부전의 다양한 병리학적 측면에 영향을 미친다(Yang et al., 2015). 유전학적 원인에 의한 심부전에서 가장 빨리 나타나는 임상적인 특징은 심실 공간의 확장 없이 심실벽이 비후해 지거나, 심실벽 두께의 변화 없이 심실의 공간이 확장되는 변화를 보이는 것이다. 각각의 심실 구조 변화는 특이적인 혈류역학적 변화를 가지고 오는데, 심실벽이 비후해지면 심장의 수축 기능에는 이상이 없지만 이완기에 이상이 생기는 반면, 심실이 확장되면 심장의 수축기능이 저하된다. 이러한 임상적 증상이 나타나면 비대성 이나 확장성 심근증으로 진단이 된다(Morita et al., 2005).

세계적으로 심부전 환자들이 증가함에 따라 이에 대한 이환율과 사망률 그리고 의료비용은 사회적 문제가 되고 있다. 미국의 경우에는 연간 50만명이 새로 진단을 받고 37억 달러의 비용이 지불되며 5만명이 사망한다(Go et al., 2013). 한국에서도 심부전 환자가 빠르게 증가하고 있는데 2002년에 약 0.75%였던 유병률이 2013년에는 약 1.53%로 증가했으며 2040년에는 약 3.35%로 증가 할 것이라 예측된다(Lee et al., 2016). 심부전 환자들의 생존율이 개선 되기는 했지만 5년 생존율은 많은 암과 비슷할 정도로 높은 편이다(Askoxylakis et al., 2010). 또한 말기 심부전 환자의 경우에는 1년 생존율이 50%정도로 낮을 뿐만 아니라 인공심장이나 심장이식 등 고비용 치료가 필요하기 때문에 심부전에 대한 조기진단의 중요성이 커지고 있다.

1.1.4 심근증 관련 데이터베이스

인간 질병에 대한 정보는 질병의 치료를 위한 분자적 기전을 밝히고 유전자 변이와 표현형의 관계를 밝히는 한편 연구 결과를 임상치료에 적용하기 위한 생물학적 연구에 꼭 필요하다(Kibbe et al, 2015). 인간 질병에 대한 정보를 효율적으로 분석하기 위해서는 데이터베이스의 구축이 중요하며, 그렇기 때문에 질병에 대한 유전 및 임상 등에 대한 정보가 수록되어 있는 데이터베이스가 많이 구축되어있다. <Table 1.8>은 심근증과 관련된 정보를 얻을 수 있는 데이터베이스들이다. 유전성 심근증은 유전학적으로 다양한 형태의 변이가 나타나고 정상인에게서는 잘 나타나지 않는 유전 변이에 의한 경우가 많기 때문에 유전자 검사를 통해 심근증의 조기 진단 및 예후의 예측, 치료의 단서를 얻을 수 있다(Golbus et al, 2014). 또한 최근에는 심근증과 관련이 있는 단백질들의 상호작용과 생체 경로에 대한 연구 결과가 늘어나고 있는데, 이를 이용한 심근증의 발병 기전의 이해를 통해 효과적인 치료법이나 치료의 타겟이 되는 단백질들을 밝혀 낼 수 있는 가능성이 제기되고 있다(Roselló-Lletí et al., 2014; Yoshikawa et al., 2009).

인간 유전질환 정보 통합 데이터베이스(Online Mendelian Inheritance in Man, OMIM)

OMIM는 전신인 Dr Victor McKusick의 Mendelian Inheritance in Man (MIM)이 1985년에 Johns Hopkins University School of Medicine (JHUSOM)의 Welch Medical Library와 Howard Hughes Medical Institute가 합작됨에 따라 온라인 사이트로 개설 되었으며, 1995년부터는 미국 국립생물정보센터(National Center for Biotechnology Information, NCBI)에 의해 관리되고 있다(Amberger et al., 2011). OMIM은 유전자의 표현형을 검색 할 수 있고, 특정 유전 질환과 관련된 유전자와 그

변이, 그 유전자들이 암호화 하고 있는 단백질 등, 유전질환에 대한 다양한 정보를 담고 있으며 연관된 다른 유전학적 데이터베이스들과도 연동되어 있다(Amberger et al., 2009). OMIM을 통해 심근증의 원인이 되는 유전자 및 단백질에 대한 정보뿐만 아니라 유전자의 변이에 따른 임상적인 증상에 대한 정보까지 얻을 수 있는데, 이러한 정보들은 심근증과 관련이 있는 단백질들간의 상호작용 가중치를 계산해서 질병을 일으키는데 우선적으로 작용하는 단백질을 알아보는 연구 등에 사용 되기도 한다(Li W et al., 2013).

GenBank 데이터 베이스

GenBank는 미국의 NCBI에서 운영하는 유전자 정보 데이터베이스이다. GenBank는 Human Genome Project 결과를 포함하여 세계 각지의 연구실에서 생성된 DNA 염기서열 정보들을 모아 놓은 것으로, 1992년 10월 처음으로 서비스를 시작하였다. 현재 GenBank는 DNA 염기서열 정보와 단백질 정보를 함께 제공하고 있다. 유전자 정보를 담고 있는 주요한 데이터베이스로는 GenBank외에도 일본의 DDBJ (DNA Data Bank of Japan)와 유럽의 EMBL (European Molecular Biology Laboratory)이 있는데, 이들은 국제 염기서열 데이터베이스 연합(International Nucleotide Sequence Database Collaboration)을 통하여 거의 동일한 정보를 공유하고 있다. GenBank 플랫폼의 주요 행의 내용은 다음과 같다. DEFINITION 행은 그 자료의 생물학적 특징을 요약하여 나타내는 행이다. ACCESSION number는 서열자료의 고유한 식별자로서, 분자 서열에 대응하는 문자와 숫자로 이루어져 있는데, 고유한 염기서열 각각에 대해 이 검색번호가 부여되며, 이것은 데이터베이스의 기록을 조회하는데 중요한 요소이다. DEFINITION 행은 그 자료의 생물학적 특징을 요약하여 나타내는 행이다. SOURCE행에는 서열자료가 얻어진 생물체의 일반명이나 과학명이 기재된다. ORGANISM에는 서열자료가 유래된 생물의 과학적 학명이 NCBI 분

류데이터베이스에서 사용하는 계통 분류에 근거하여 표시된다. ORIGIN은 서열이 유래한 원래 source의 전체 서열을 나타낸다 (Benson et al., 2013).

ClinVar 데이터 베이스

ClinVar는 미국의 NCBI에 의해 2013년에 구축되었는데, 임상적으로 중요한 유전자 변이와 표현형에 대한 정보를 담고 있다. ClinVar는 dbSNP 및 dbVar와 연동되어 있는데 이는 인간의 유전체에서 유전 변이의 위치를 다루고 있는 데이터 베이스들이다. 또한 의학 유전학적인 정보를 담고 있는 Medical Genetics 데이터베이스와 표현형에 대한 정보를 공유 하고 있다(Landrum et al., 2014). ClinVar는 표현형에 따른 유전자 변이뿐 만 아니라 clinical significance항목을 통해 해당 변이가 표현형에 미치는 임상적 의의를 확인할 수 있다. Clinical significance의 정도는 ‘Pathogenic’, ‘Likely pathogenic’, ‘Likely benign’, ‘Benign’으로 나타내며 정보를 제공하는 사람들간에 임상적 의의가 서로 다르면 ‘Conflicting interpretations of pathogenicity’로 나타내고 때에 따라서 clinical significance항목을 제공 하지 않는 경우도 있다(Harrison et al., 2016).

보편적 단백질자원 데이터베이스(universal protein resource, UniProt)

UniProt은 2002년 영국의 the European Bioinformatics Institute (EBI), 스위스의 the Swiss Institute of Bioinformatics (SIB), 그리고 미국의 the Protein Information Resource (PIR)의 합작으로 구축 되었다. UniProt은 단백질의 서열과 주석(annotation)에 대한 대량의 정보를 보유하고 있으며 단백질의 아미노산 서열의 유사도를 검색 할 수 있는 ‘basic local alignment search tool (BLAST)’, 여러 서열을 정렬 할 수 있는 ‘Align’, 외부 데이터베이스의 정보를 불러올 수 있는 ‘Retrieve/ID Mapping’과 같은 생물정보학적 분석도구들을 제공 하고

있다(Pundir et al., 2002). UniProt을 통해 심근증과 관련이 있는 단백질을 확인 할 수 있고, 해당 단백질의 기능, 단백질의 기능을 변화시키는 유전변이, 유전자 온톨로지(gene ontology)주석, 아미노산의 서열 등에 대한 정보를 얻을 수 있다(Frazier et al., 2008).

부정맥 유발성 우심실 심근증 유전 변이 데이터베이스 (ARVD/C genetic variants database)

ARVD/C 데이터베이스는 2008년에 네덜란드(Netherlands)의 Department of Genetics of the University Medical Center Groningen에 의해 구축 되었다. ARVD/C 데이터베이스는 부정맥 유발성 우심실 심근증과 관련이 있는 12개의 유전자와 이 유전자의 약 1400여개의 변이에 대한 정보를 담고 있는데, 총 변이 중 411개의 변이만 임상적 의의가 알려져 있고 나머지 변이들을 아직 확실히 알려져 있지 않다. ARVD/C 데이터베이스에서 제공하는 정보에는 유전자와 유전자좌위 (gene and locus), 엑손과 인트론의 경계부위(exon-intron boundaries), 표준서열(reference sequence)에서의 DNA염기 변화, 단백질 변화, 변이의 종류, 병원성 환경과 변이에 대한 연구 결과들이 포함되어있다. ARVD/C 데이터베이스는 지금까지 알려진 모든 부정맥 유발성 우심실 심근증 관련 유전정보들을 개략적으로 제공하고 있으며, 해당 질병에 대한 가장 광범위한 정보를 담고 있다(Lazzarini et al., 2015).

Table 1.8 Database deal with information of cardiomyopathy

Database	URL	Year of construction	Explain
Online Mendelian inheritance in man (OMIM)	http://www.ncbi.nlm.nih.gov/omim	1985	Information on genes, proteins and phenotypes related to genetic diseases
GenBank	www.ncbi.nlm.nih.gov/genbank	1992	DNA sequence information generated in the laboratory
ClinVar	www.ncbi.nlm.nih.gov/clinvar	2013	Information on clinically important gene mutations and phenotypes
Universal protein resource (UniProt)	http://www.uniprot.org	2002	Information on sequence and annotation of proteins. Sequence similarity, sequence alignment, data retrieval is possible.
ARVD/C genetic variants database	http://www.arvcdatabase.info	2008	Information on 12 genes and 1,400 mutations associated with ARVD / C

There are varieties of databases abroad based on diverse genetic information about cardiomyopathy. But There is no database specific to cardiomyopathy except ARVD/C genetic variants database.

1.1.5 마이크로어레이와 차별발현 유전자 분석

유전자 발현에 대한 연구의 목적 중 하나는 특정 상황 하에서 상황 혹은 하향 조절되는 유전자를 찾는 것인데, 이는 실험 과정에서 많은 양의 데이터가 생성되기 때문에 컴퓨터를 이용한 데이터 저장 및 분석이 요구 된다(Pasternak J, 2005). DNA 마이크로어레이는 수만 개의 유전자의 발현을 동시에 정량적으로 측정 할 수 있는 대량고속처리 기술로 서로 다른 장기들 혹은 개체들 간의 유전자 발현을 비교하여 새로운 유전자를 찾거나 유전역학적으로 의미가 있는 생체지표를 발견하는데 이용되고 있다(Suárez et al., 2007).

마이크로어레이를 통한 유전자 발현 정도를 분석할 때는 여러 실험들 간의 결과 차이를 보정하기 위해 background correction, data normalization, data summarization 등의 데이터 전처리가 필요하다. 마이크로어레이는 광학 스캐너를 이용하여 발현 강도를 측정하기 때문에 비특이적 반응을 보정 해 주어야 하는데, 이 과정을 background correction 이라 한다(Verducci et al., 2006). 유전자 발현 정도를 분석 할 때는 여러 마이크로어레이 실험의 결과를 조합하여 분석 하는데, 같은 조건으로 실험을 진행 한다고 해도 각 실험의 환경이나 기기적 조건에 의하여 실험간 편향이 생길 수 있다. 이러한 편향을 상쇄시키기 위해 data normalization 과정을 거친다. 많은 방법들 중 스케일링 기법은 Affymetrix의 표준 normalization 방법인데 먼저 기준 실험의 2% 절삭평균을 구하고 이 값을 각 실험의 2% 절삭평균으로 나누어 기준실험에 대한 각 실험의 보정계수를 구한다. 이 보정 계수를 각각 해당하는 실험의 결과에 곱해주면 각 실험데이터가 기준 실험을 바탕으로 표준화되어 실험간 편향이 상쇄된다(Bolstad et al., 2003). Normalization을 한 뒤에는 여러 probe의 발현 정도를 하나로 맞춰주는 과정인 summarization을 하는데, 이 단계가

지 하면 마이크로어레이 분석을 위한 데이터 전처리 과정이 끝난다.

데이터의 전처리가 끝나면 통계분석을 하게 되는데 다양한 조건에서 차별적으로 발현이 조절되는 유전자를 찾기 위해서는 t-test 혹은 ANOVA가 주로 사용된다. 이 중 두 그룹에 대한 짝 비교에는 t-test가 사용된다. 통계분석 방법 중 t-test를 사용하기 위해 발현 데이터를 전처리 한 뒤 \log_2 스케일로 변환하여 데이터들이 t-test를 하기에 적당한 형태로 분포하게 한 뒤 통계분석을 실시한다(Chiaretti et al., 2004). 마이크로어레이를 통해 차별 발현되는 유전자에 대해 연구할 때는 많은 수의 유전자를 비교 대상으로 하기 때문에 다중가설검정의 문제가 생기게 된다. 즉 검정통계량으로 p-value를 사용할 경우 위양성이 많아지게 된다. 이를 보정하기 위해 family wise error rate (FWER)나 false discovery rate (FDR)보정 방식을 사용하게 되는데, FWER 방식은 너무 과도하게 보정하여 발현량이 유의하게 차이나는 유전자를 너무 적게 얻을 수 있기 때문에 마이크로어레이에서는 주로 FDR이 사용되고 FDR의 %를 다양하게 조절하여 발현량에 유의한 차이를 보이는 유전자의 수를 조절한다(Verducci et al., 2006).

위의 방법으로 실험군과 대조군 사이에 발현량이 유의하게 차이가 나는 유전자들이 정해지면 해당 유전자들의 의미론적 해석을 위한 기능과 생물학적 경로를 알아보기 위해 Gene ontology (GO)와 KEGG database 기반 pathway 분석을 하게 된다. GO는 cellular component, molecular function, biological process, 이렇게 3가지 범주의 주석 정보를 제공하는데 이를 통해 차별발현이나 클러스터 분석 등에서 얻어진 유전자 목록의 의미론적 공통점을 찾을 수 있다. 유사한 GO 주석을 소유한 유전자는 기능적으로도 서로 유사한 유전자일 것이라는 가정을 할 수 있다. 그러므로 유전자에 주석된 GO 용어는 유전자와 유전자 사이 혹은 유전자 클러스터와 클러스터 사이의 의미론적 유사성 평가에 이용 될 수 있다..

1.1.6 유전자 네트워크 분석

특정 유전자의 발현량 변화가 다른 유전자의 발현에 미치는 영향은 상관관계를 포함한 다양한 통계적 방법으로 연구 할 수 있는데, 이를 이용하여 유전자들간의 상호작용을 알아 볼 수 있다. 이를 통해 유전자들의 네트워크를 구성 할 수 있는데, 유전자 네트워크에서는 유전자 또는 유전자산물을 노드로, 분자간 상호작용을 연결선으로 표현하고 이들의 상호작용 집합체로 네트워크를 구성한다.

유전자 네트워크를 분석하기 위한 도구 중의 하나는 시스템 유전학을 위한 웹 서비스인 GeneNetwork (www.genenetwork.org, GN)인데 이는 다양한 분자적, 표현형에 대한 데이터들을 통합시킨 데이터베이스인 동시에 분석 도구이다(Mulligan et al., 2016). GN의 인간에 대한 대부분의 데이터 세트는 유전자 발현량을 포함 하고 있으며 the Genotype-Tissue Expression (GTEx)프로젝트의 데이터 세트와 같은 몇몇의 데이터 세트는 양적 형질 유전자좌(quantitative trait loci)분석에 이용 될 수 있는 데이터를 포함 하고 있다. 이러한 데이터 세트의 타입은 검색 결과에 영향을 주는데 유전자와 단백질 발현 데이터 세트는 발현 정도, 유전자의 위치, 유전자 기호, Gene Ontology 카테고리 등에 기반하여 결과를 도출한다. 이를 통하여 유전자들의 발현에 대한 상관분석을 할 수 있는데 여기에는 피어슨 상관관계, 스 피어만 상관관계 등이 이용된다. GN에서는 분석 하고자 하는 유전자들 간의 발현 상관관계 행렬을 얻을 수 있으며 이를 이용하여 유전자들간의 네트워크를 구성 할 수 있다.

GN에도 네트워크를 구현해 주지만 유전자들 사이의 상관성 데이터를 바탕으로 R 프로그래밍 언어(R)등을 이용하여 소셜네트워크 분석에 기초한 유전자 네트워크 분석을 수행 할 수 있다. 소셜네트워크는 개인과 집단, 사회의 관계를 네트워크로 파악하는 개념인

데, 네트워크에서 개인 또는 집단, 사회가 하나의 노드이며 소셜네트워크는 각 노드들 간의 상호의존적인 관계에 의해 만들어지는 사회적 관계 구조이다. 모든 노드들은 네트워크 안에 존재하는 개별적인 독립체들이고 연결선은 각 노드들 간의 관계를 나타내는데, 소셜네트워크적인 관점은 각각의 독립체들이 이루는 구조에 대한 일련의 연구 방법들을 제시해 준다(Wasserman and Faust, 1994). 소셜네트워크분석은 다수의 노드와 이들을 연결하는 연결선으로 구성된 구조를 네트워크 이론과 그래프 이론을 통해 분석하는 과정이다(Otte and Rousseau, 2002). 소셜네트워크의 분포에 대한 분석 중에는 중심성에 대한 분석이 있는데 이는 네트워크에서 특정 노드의 중요도나 영향력을 정량적으로 평가하는 지표가 된다. 중심성을 측정하는 대표적인 방법으로는 연결정도 중심성, 근접중심성, 중개중심성 등이 있다(Newman, 2010). 연결정도 중심성은 네트워크에서 한 노드에 다른 노드들이 연결되어 있는 정도 말하는데, 이는 다시 연결방향의 비율에 따라 노드로 오는 방향의 연결 정도를 측정하는 내향중심성과 노드에서 밖으로 나가는 방향의 연결 정도를 측정하는 외향중심성이 있다. 근접중심성은 한 노드가 다른 노드들에 얼마나 가깝게 위치 하는가를 말하는 개념으로 네트워크에서 각각의 노드로부터 다른 모든 노드로의 경로거리에 초점을 맞춘 지표이다. 중개중심성은 네트워크에서 하나의 노드가 다른 노드들 사이에 위치하는 정도를 나타내는 지표로 중개중심성이 크다는 것은 그만큼 많은 노드의 사이에 위치하고 있다고 볼 수 있고, 2개의 노드를 연결하는 커뮤니케이터의 역할, 혹은 허브의 역할을 한다(Borgatti and Everett, 2006). 유전자 발현 정보를 통해 구성 된 유전자 네트워크의 구조를 연구하기 위해서 소셜네트워크 분석법을 사용 할 수 있는데, 네트워크의 중심성을 지표로 하여 핵심 유전자들과 생물학적 시스템을 조절하는데 중요한 허브 유전자들을 찾을 수 있다(Ohn et al., 2003).

1.2 연구의 필요성

관상동맥성심질환(coronary heart disease, CHD)이 주로 노인들에서 많이 발병 하는 것 과는 다르게 심근증은 젊은 연령대에서도 발병 할 수 있다. 즉, 모든 연령대에서 심근증으로 인한 심부전, 부정맥, 혈전증(thromboembolism)등의 증상이 나타날 수 있다. 심근증으로 인한 심장의 기능장애는 평생 동안 증상이 나타나지 않는 것에서부터 심부전, 부정맥, 혈전증, 심장돌연사 등에 이르기까지 매우 다양하다. 심근증을 치료하는 방법은 심근증의 종류, 증상의 정도에 따라 매우 다양한데, 단순히 생활 습관을 바꾸는 것, 혈압을 낮춰주거나 심박수 저하, 혈전이 생기는 것을 방지, 체내 전해질 균형 유지 등을 위해 다양한 약물을 사용하는 방법, 격벽 근육절제를 위한 외과수술적 방법, 심장의 리듬을 조절해 주거나 기능을 향상시켜 줄 수 있는 장비를 삽입하는 방법, 심장 이식수술 등이 있다(Maron et al., 2013). 위와 같이 심근증은 증상과 치료법이 매우 다양하기 때문에 효율적인 치료를 위해서는 조기에 정확하게 진단을 하는 것이 중요하다.

최근 염기서열 분석법의 발달로 심근증과 관련이 있는 유전자와 각 유전자의 변이가 밝혀 지고 있다. 또한 각 유전자의 변이가 임상적으로 어떠한 영향을 끼치는지에 대한 정보도 늘어나고 있는데, 이러한 근거를 가지고 심근증에 대한 유전학적 진단검사를 시행하는 기관이 늘고 있다. NGS등의 발전된 염기서열 분석법들은 기존의 생어(Sanger Sequencing)방법에 비해 훨씬 많은 유전자를 빠른 시간에 분석 할 수 있게 해 주는데, 이는 질병과 관련된 전체적인 유전자 분석 결과를 제공하여 심근증을 진단 하는데 도움을 주고 있다(Limongelli et al, 2012). 심근증은 별다른 증상을 보이지 않거나, 증상을 표현하기 힘든 유아들에게서 심장돌연사를 일으킬 수 있으며,

이는 소아들의 급사(sudden death)에 주요한 요인이 되기도 한다(Ulla et al., 2016). 이러한 경우에는 심근증에 대한 유전자 검사가 증상이 나타난 후 시행하는 고전적인 진단법들을 보완해주어 조기에 심근증을 발견 하고 치료 방법을 결정하는데 도움을 줄 수 있다.

이렇게 심근증에 대한 유전학적 연구가 늘어남에 따라 공용 데이터 베이스를 통해 얻을 수 있는 유전체, 단백질체, 임상적 정보가 많아지고 있지만 심근증에 대한 유전학적 데이터베이스는 2008년 네덜란드에서 구축 된 부정맥 유발성 우심실 심근증 유전 변이 데이터베이스가 유일하다. 이에 따라 임상적으로 유용하게 활용 될 수 있는 심근증에 대한 통합적인 데이터베이스의 구축과 이를 통한 효율적인 유전학적 정보제공의 필요성이 있다.

또한 마이크로어레이 같은 대량 신속처리 기술의 발전으로 수만 개의 유전자의 발현 정도를 한번에 정량적으로 측정 할 수 있게 됨에 따라 기존의 전통적인 분자생물학적 연구방법의 단일 유전자를 대상으로 연구를 진행할 수 밖에 없었던 한계를 넘어 수많은 유전자들 사이의 상호작용을 밝혀 낼 수 있게 되었다. 최근 심근증에 대한 마이크로 어레이 분석의 증가로 유전자 발현수준의 정보들이 늘어나고 있고, 이를 이용하여 분자수준에서의 질병의 기전 및 생물학적 경로를 조절하는 유전자에 대한 연구결과들이 많이 나오고 있다. 하지만 마이크로 어레이를 이용한 유전자 발현 수준의 연구를 통해서만 주로 체세포 변이 혹은 국소적인 유전자 발현량에 대한 결과만을 얻을 수 있기 때문에 말초혈액으로 수행되는 통상적인 진단검사에 위의 연구 결과들을 적용하는 데는 한계가 있는 현실이다. 따라서 심근증의 조기 진단 및 예후의 예측을 위해 유전자 발현수준의 연구 결과들을 진단검사에 적용 할 수 있는 많은 연구가 진행되어야 할 필요가 있다.

1.3 연구의 목적

심근증에 대한 통합적인 유전학적 정보를 다루는 데이터베이스의 구축 및 유전자 발현수준의 연구결과들은 심근증에 대한 유전학적 정보를 수집하여 질병에 대한 연구를 하는데 뿐만 아니라 질병의 진단 및 예후의 예측을 위하여 유전학적 정보를 활용 하는데 기반이 될 것이다. 즉, 본 연구에서는 심근증에 대한 유전자, 단백질체, 임상적 정보를 통합한 데이터 베이스를 구축하여 심근증의 연구 및 임상에서 이용 될 수 있는 정보를 제공하고, 유전자 발현 수준에 대한 연구 결과들을 통상적인 진단검사에 적용 할 수 있는 방법에 대하여 연구하고자 하였다.

본 연구에서는 심근증과 관련이 있는 유전자의 서열 및 해당 유전자가 암호화 하고 있는 단백질의 아미노산 서열과 그것이 이루는 구조, 유전자의 변이가 가지고 있는 임상적 의의에 대한 정보를 제공하는 데이터베이스를 구축 하였다. 이는 전산학적 기술을 이용하여 심근증에 관련된 정보들을 핸들링하고 이에 기반한 데이터베이스를 웹 인터페이스를 통해 공개 함으로서, 심근증에 대한 유전학적 정보의 검색과 획득에 용이하도록 하였다. 이를 통해 유전자 각각의 변이가 발현되는 단백질의 구조적으로 어떠한 영향을 미치며, 이러한 영향이 심근증의 표현형에 어떠한 임상적 의의가 있는지 확인 할 수 있다.

또한 유전자 발현에 대한 정보를 진단검사에 적용 할 수 있는 방법을 모색하였다. 먼저 마이크로어레이를 통해 특정 표현형과 관련이 있는 유전자 목록을 얻어 유전자들 사이의 발현 상관성을 기초로 유전자 네트워크를 구성하고 여기에 선천적으로 심근증과 관련이 있는 유전자들과 2부 네트워크를 구성한다. 이렇게 구성된 2부 유전자 네트워크에 소셜네트워크 분석법을 적용하면 처음 구성된

유전자 발현 네트워크와 선천적으로 변이를 보이는 유전자들 사이의 상관 관계를 알 수 있다. 본 연구는 다양한 표현형과 원인의 심근증 환자들 중 심부전의 예후를 보이는 사람들의 유전자 발현 정보를 수집하여 심부전에서 발현량에 차이를 보이는 유전자들의 네트워크를 구성하고 여기에 상관성이 있는 선천적 변이의 유전자들을 탐색하여 말초혈액을 통한 진단검사로 심근증 환자들의 예후 중 심부전을 예측 할 수 있는 유전자 목록을 얻는데 목적이 있다.

이와 같은 심근증의 유전학적 정보에 특화된 데이터베이스 구축이 NGS등 염기서열 분석법의 발전으로 나날이 증가하는 심근증과 관련된 유전학적 정보를 수집하고 가공하여 유전학적 정보들이 심근증에 대한 연구 및 임상적 진단에 쉽게 이용 될 수 있게 하는데 이바지할 것으로 기대하며, 심근증 뿐만이 아닌 유전학적인 원인으로 발병되는 다른 질환에도 유전학적 정보 데이터베이스들이 구축되어 해당 질병에 대한 연구 및 임상적 진단에 적용 될 수 있기를 바란다.

제 2 장. 연구 방법

2.1 데이터 수집 및 가공

본 연구에서 이용된 심근증과 관련된 데이터는 GenBank (www.ncbi.nlm.nih.gov/genbank/), ClinVar (www.ncbi.nlm.nih.gov/clinvar/), UniProt (www.uniprot.org/)에서 수집하였다. GenBank에서는 유전자에 대한 염기 및 아미노산서열 등에 대한 정보를, ClinVar에서는 유전자에 대한 변이와 이와 연관된 표현형 및 임상적 의의에 대한 정보를, UniProt에서는 해당 유전자가 발현하는 단백질에 대한 정보를 얻을 수 있다. 각각의 데이터 베이스에서 ‘cardiomyopathy’를 키워드로 검색을 하였고, GenBank 와 UniProt에서는 ‘organisms’를 ‘Homo sapiens’로 지정하여 플랫폼파일을 다운받았다. 여기에서 얻은 정보들은 JAVA 프로그래밍 언어를 이용하여 데이터 파싱 과정을 수행하였다.

GenBank의 플랫폼 파일에서는 ‘ACCESSION’행의 서열 고유번호, ‘GI (GenInfo Identifier)’행의 유전자정보 검색번호, ‘DEFINITION’행의 생물학적 특징 요약, ‘CDS (coding sequence)’ - ‘translation’행의 아미노산서열, ‘ORIGIN’행의 염기서열 정보만 추출 하였다. ClinVar의 플랫폼 파일에서는 ‘Name’ 행의 변이 종류, ‘Gene(s)’ 행의 유전자 이름, ‘Condition(s)’행의 표현형, ‘Clinical significance’행의 임상적 의의에 대한 정보만 추출하였다. UniProt의 플랫폼 파일에서는 유전자 이름, 단백질 이름, 분자량, KEGG 검색 키워드, 단백질의 구조와 관련된 정보만 추출하였다. 이렇게 추출한 정보들을 sql포맷의 파일로 재 가공하였고 <Figure 2.1>, <Figure 2.2>, <Figure 2.3>를 통해 각각의 플랫폼파일과 재 가공 된 sql포맷 파일의 내용을 확인 할 수 있다.

a)

```

LOCUS       S80782                69 bp    mRNA    linear    FRI 14-JUL-2016
DEFINITION  MyBP-C=cardiac myosin binding protein-C (alternatively spliced)
            [human, familial hypertrophic cardiomyopathy patient,
            lymphoblastoid cell lines, mRNA Partial Mutant, 69 nt].
ACCESSION   S80782
VERSION     S80782.1  GI:1245895
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69)
  AUTHORS   Bonne,G., Carrier,L., Bercovici,J., Cruaud,C., Richard,P.,
            Hainque,B., Gautel,M., Labeit,S., James,M., Beckmann,J.,
            Weissenbach,J., Vosberg,H.P., Fiszman,M., Komajda,M. and
  TITLE     Cardiac myosin binding protein-C gene splice acceptor site mutation
            is associated with familial hypertrophic cardiomyopathy
  JOURNAL   Nat. Genet. 11 (4), 438-440 (1995)
  PUBMED    7493026
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 175042] from the original journal article.
FEATURES             Location/Qualifiers
     source          1..69
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..69
                     /gene="MyBP-C"
                     /note="cardiac myosin binding protein-C"
     CDS             1..69
                     /gene="MyBP-C"
                     /codon_start=1
                     /product="cardiac myosin binding protein-C"
                     /protein_id="AA35663.1"
                     /db_xref="GI:1245895"
                     /translation="FRQDPPLPRPHTRHHCSCSWK"
ORIGIN
     1 cccaggcagg atccacctgg actgccagg ccgcatacca gacaccattg tggttgtagc
    61 tggaaataa
//

```

[illegible]

a)

Name	Gene(s)	Condition(s)	Frequency	Clinical significance	(Last reviewed)	Review status	Chromosome	Location	Assembly	VariationID	AlleleID
PM_000745.1	AARS2	c.64:G>A	cy- <i>Cypr15</i> (Leufs)	AARS2 Combined oxidative phosphorylation deficiency 8	Pethogenic(Last reviewed: Jul 2, 2014)						
PM_000745.1	AARS2	c.177C>T	AARS2	Combined oxidative phosphorylation deficiency 8 (not pathogenic)(Last reviewed: Mar 29, 2016)	criteria provided, single submitter						
PM_000693.1	ACMGF1	c.335T>C	Likely benign(Last reviewed: Mar 29, 2016)	criteria provided, single submitter							
PM_000693.1	ACMGF1	c.335G>A	Likely benign(Last reviewed: Mar 29, 2016)	criteria provided, single submitter							
PM_000693.1	ACMGF9	c.305A>G	G_Sln13(TrnMet)	Uncertain significance(Last reviewed: Jan 26, 2014)	criteria provided, single submitter						
PM_000693.1	ACMGF9	c.243C>T	ACMGF9 Dilated cardiomyopathy 10	Uncertain significance(Last reviewed: Feb 26, 2014)	criteria provided, single submitter						
PM_000693.1	ACMGF9	c.1975A>G	c.TheT59Ala	ACMGF9 Dilated cardiomyopathy 10	Uncertain significance(Last reviewed: Feb 26, 2014)	criteria provided, single submitter					
PM_000693.1	ACMGF9	c.1165>ZD	ACMGF9 Dilated cardiomyopathy 10	Uncertain significance(Last reviewed: Jan 16, 2014)	criteria provided, single submitter						
PM_000693.1	ACMGF9	c.1012>D	ACMGF9 Dilated cardiomyopathy 10	Uncertain significance(Last reviewed: Mar 11, 2014)	criteria provided, single submitter						
PM_000693.1	ACMGF9	c.1043T>G	c.LeuS48Ala	ACMGF9 not specified	Uncertain significance(Last reviewed: Oct 16, 2013)	no assertion criteria provided					
PM_000693.1	ACMGF9	c.459C>T	c.ProI51Ser	ACMGF9 not specified	Uncertain significance(Last reviewed: Sep 22, 2015)	criteria provided, single submitter					
PM_000693.1	ACMGF9	c.4519G>A	c.ValI150Tleu	ACMGF9 not specified	Uncertain significance(Last reviewed: Mar 4, 2015)	criteria provided, single submitter					
PM_002097.1	ACMGF9	c.450C	ACMGF9 not specified	Uncertain significance(Last reviewed: Aug 17, 2015)	criteria provided, single submitter						

```

insert into cardio_clinVar
values ('NM_005691.3(ABCC9):c.3316-4A>C', 'ABCC9', 'DCM', 'Likely benign', 'intron variant');
insert into cardio_clinVar
values ('NM_005691.3(ABCC9):c.3095A>G (p.GLM1032ARG)', 'ABCC9', 'DCM', 'Uncertain significance', 'single nucleotide variant');
insert into cardio_clinVar
values ('NM_005691.3(ABCC9):c.2630C>T (p.THR877MET)', 'ABCC9', 'DCM', 'Uncertain significance', 'single nucleotide variant');
insert into cardio_clinVar
values ('NM_005691.3(ABCC9):c.1975A>G (p.THR659ALA)', 'ABCC9', 'DCM', 'Uncertain significance', 'single nucleotide variant');
insert into cardio_clinVar
values ('NM_005691.3(ABCC9):c.1165-2A>G', 'ABCC9', 'DCM', 'Uncertain significance', 'splice acceptor variant');
insert into cardio_clinVar
values ('NM_005691.3(ABCC9):c.1012-1G>C', 'ABCC9', 'DCM', 'Uncertain significance', 'splice acceptor variant');
insert into cardio_clinVar
values ('NM_005691.3(ABCC9):c.4570_4572DELTTAAAGAAAT (p.LEU1524LYFS)', 'ABCC9', 'DCM', 'Uncertain significance', 'indel frameshift');

```

a)

```

ID   MYH7_HUMAN                               Reviewed:      1935 AA.
AC   P12883; A2TDB6; B6D424; Q14836; Q14837; Q14904; Q16579; Q2M1Y6;
AC   Q92679; Q9H1D5; Q9UDA2; Q9URM8;
DT   01-OCT-1989, integrated into UniProtKB/Swiss-Prot.
DT   06-DEC-2005, sequence version 5.
DT   06-JUL-2016, entry version 199.
DE   RecName: Full=Myosin-7;

FT   DOMAIN            85   778   Myosin motor.
FT   DOMAIN            781   810   IQ. (ECO:0000255|PROSITE-
FT                                     ProRule:PRU00116).
FT   NP_BIND           178   185   ATP.
FT   REGION            655   677   Actin-binding.
FT   REGION            757   771   Actin-binding.
FT   COILED            839   1935   [ECO:0000255].
FT   MOD_RES           129   129   N6,N6,N6-trimethyllysine. (ECO:0000255).

FT   HELIX             1590  1655   [ECO:0000244|PDB:5CHX].
FT   HELIX             1777  1806   [ECO:0000244|PDB:4XA6].
FT   HELIX             1810  1841   [ECO:0000244|PDB:4XA6].
FT   TURN              1842  1846   [ECO:0000244|PDB:4XA6].
FT   HELIX             1847  1856   [ECO:0000244|PDB:4XA6].
FT   HELIX             1864  1883   [ECO:0000244|PDB:4XA6].
FT   HELIX             1891  1899   [ECO:0000244|PDB:4XA6].
SQ   SEQUENCE   1935 AA;  223097 MW;  C58B22F914215718 CRC64;
MDSSEGVVFG AAPPYLKGE KERLEAQTTP FDLGGVTVF DQGEFVQAK IVSREGGKVT
AEIYQKIVT VKEDQVMQCN PPKFDKIDM AMLTFLNEPA VLYNLKDRYG SWMIYTSGL
FCVTVPVQKM LFCOTPEVVA AVGKQKSEA PPHIFSISDN AVQVGLTDFE NOSILITGES

VEAVQECRN AEKAGGATY DAHDAELK KQQTSAHLK KKKKNGQTI KDLQRLEDA
EQIALKGGKK QLQKLEARVR ELENELEAEQ KRNAESVQGM RKSERRIKEL TYQTEEDRGN
LLRLQLVDK LQLKVGAYKR QAEEAEQAN TNLKTRKQV HELDEAEKRA DIAESQVWKL
RAKSRDIGTK GLNEE
//

```

b)

```

values ('P00156', 'Cytochrome b', 'MT-CYB', 'hse:4519', '380', '42718', '', '', 'METAL', 83, 83, Iron 1 (hes
METAL 97 97 Iron 2 (heme b566 axial ligand)
METAL 182 182 Iron 1 (heme b562 axial ligand)
METAL 196 196 Iron 2 (heme b566 axial ligand)', 'CHAIN 1 380 Cytochrome b', 'TRANSMEM
TRANSMEM 76 98 Helical
TRANSMEM 115 135 Helical
TRANSMEM 138 158 Helical
TRANSMEM 178 198 Helical
TRANSMEM 230 250 Helical
TRANSMEM 288 308 Helical
TRANSMEM 323 343 Helical
TRANSMEM 349 369 Helical');
insert into cardio_uniProt
values ('P13533', 'Myosin-6', 'MYH6', 'hse:4624', '1939', '223735', '', 'DOMAIN', 85, 780, Myosin motor
DOMAIN 783 812 IQ
REGION 657 679 Actin-binding
REGION 759 773 Actin-binding
COILED 842 1939', 'NP_BIND 178 185 ATP', 'CHAIN 1 1939 Myosin-6
MOD_RES 129 129 N6,N6,N6-trimethyllysine
MOD_RES 379 379 Phosphothreonine
MOD_RES 417 417 Phosphoserine
MOD_RES 1139 1139 Phosphoserine
MOD_RES 1261 1261 Phosphotyrosine
MOD_RES 1271 1271 Phosphoserine
MOD_RES 1273 1273 Phosphothreonine

```

Figure 2.3 UniProt flat file and sql file format. a) text file format from UniProt, it contains much information for example protein structure. b) sql format for table which has been regulated by JAVA programming language.

2.2 시스템 개발 환경 및 데이터베이스 구현

심근증에 대한 정보 검색 시스템 구축을 위해 Intel(R) Core(TM) i5-3427U 1.40GHz, 8Gb RAM, 238Gb SSD 사양의 서버를 구축하였다. Operating System은 Window 7 Home Premium K를 사용했고 데이터베이스 구축을 위해 MySQL 서버 5.7, 웹 서버 프로그램으로 Apache-tomcat 8.0을 설치하였다. 또한 데이터 파싱을 위해서는 JAVA 프로그래밍 언어를, 웹과의 연동을 위한 프로그래밍 언어로는 JSP, HTML 그리고 JAVA 스크립트를 사용하였다.

웹 서버를 이용해서 여러 데이터베이스에 수록된 심근증에 대한 정보들을 통합적으로 보여주고 목적에 따라 다룰 수 있게 하기 위해서 데이터베이스의 구현은 중요한 과정이다. 본 연구에서는 심근증에 대한 통합적인 유전학적 정보를 제공하고 2차 연구를 위한 염기 및 아미노산의 서열을 fasta형식의 파일로 출력하는데 그 목적이 있다. 이를 위하여 데이터베이스 내에 ‘cardio_GB’, ‘cardio_clinVar’, ‘cardio_uniProt’ 이렇게 세 개의 테이블을 만들고 GenBank, ClinVar, UniProt에서 얻은 정보를 각각의 테이블에 저장한 다음 세 개의 테이블이 유기적으로 연동되어 정보들이 출력되도록 하였다. 테이블에 저장되어 있는 데이터는 정수형과 문자형이며, 테이블 명과 테이블에 따른 필드명, 각각의 데이터 타입 등 테이블의 구조는 <Table 2.1>를 통해 확인 할 수 있다.

본 데이터베이스의 구현은 크게 두 가지로 나눌 수 있다. 첫 번째는 심근증에 대한 통합적인 유전학적 정보의 제공이다. 심근증의 분류에 따라 영향을 미치는 유전자를 검색 할 수 있고, 이렇게 검색된 유전자가 발현하는 단백질의 구조 및 각각의 유전자 변이에 대한 정보를 임상적 의의 및 표현형의 종류에 따라 분류하여 출력되게 하였다. 이는 염기 및 아미노산의 서열 정보, 유전 변이와 변

이에 따른 표현형 및 임상적 의의, 발현되는 단백질의 구조 등 여러 데이터베이스에 산재한 정보들을 통합적으로 제공하여 심근증의 유전학적 연구 및 진단에 도움을 줄 수 있을 것이다. 두 번째는 2차 연구를 위한 염기 및 아미노산 서열을 fasta 형식으로 출력하는 것이다. 기호 ‘>’로 시작하여 GI-ACCESSION-DEFINITION-sequence (DNA or amino acid)의 항목을 나타내는 fasta 형식의 데이터는 유전자 이름으로 조회 할 수 있도록 하였으며 화면으로 출력하거나 확장자가 ‘.fasta’인 파일로 다운로드 할 수 있도록 하였다. <Figure 2.4>를 통해 자세한 형식을 확인 할 수 있다.

a)

```
>gi|544063467|NM_001282626|Homo sapiens lamin A/C (LMNA), transcript variant 7, mRNA.
aggaggacctaattagagaccccttgcggcggtcggtgactcagtggttcgggggagcgccgcacctacacc
agcccaaccagatcccgagggtccgacagcgcccgccagatccccacgcctgccaggagcaagccgaga
gccagccggcgccgactccgactccgagcagctctctgtccttcgaccccgagccccgcgcccctttccgg
gacccctgcggcgccgagcgctgccaacctgcggccatggagaccccgctcccagcgcgccgcccacc
gcagcgggcgccagcgccagctccactccgctgtcgccccacccgcacacccggctgcaggagaaggagga
cctgcaggagctcaatgatcgcttggcggtctacatcgaccgtgtgcgctcgctggaaacggagaaacgca
gggctgcgccccttcgcatcaccgagctctgaagagggtggtcagccgaggtgtccggcatcaaggccgccc
acgaggccgagctcggggatgcccgcaagacccttgactcagtagccaaggagcgcccgccctgcagct
ggagctgagcaaatgcgtgaggagtttaaggagctgaaagcgcgcaataccaagaaggaggggtgacctg
atagctgctcaggctcgggctgaaggacctggaggctctgctgaaactccaaggaggccgactgagcactg
ctctcagtgagaagcgccagctggaggcgagctgcatgatctgccccggcagggtggccaagcttgagggc
agccctagggtgagggccaagaagcaacttcaggatgagatgctgcggcggggtggatgctgagaacaggctg
cagaccatgaaggaggaaactggactccagaagaacatctacagtgaggagctgcgtgagaccaaagcgcc
gtcatgagaccccgactgggtggagattgacaatgggaagcagcggtgagtttgagagccggctggcggaatgc
gctgcaggaaactcggggccccagcatgaggacacagggtggagcagttataagaaggagctgggagaagacttat
tctgccaaagctggacaatgccaggcagctctgctgagagggaacagcaacctgggtgggggctgcccacgagg
agctgcagcagctcgcgcatccgcatcgacagccctctctgccagctcagccagctccagaagcagctggc
agccaaggaggcgaaacttcgagacctggaggactcactggcccgtagcggggacacccagccggcggtg
ctggcggaagaaaggagcggggagatggccgagatgccccgaaggatgcagcagcagctggacgagtagaccag
agcttctggacatcaagctggccctggacatggagatccacgcctaccgcaagctcttggaggcgaggga
ggagaggtcactcgcttccccccagccctacctcgagcgagccgtggccgtgcttctctcactcatcc
cagacacaggggtggggcgagcgtcaccaaaaagcgcaaaactggagtcactgagagccgcagcagcttct
gaggacagcttgagccgggccccctgggcttggccgtgctgtgattccactacacctggctgaggttctctct
gcttgcggcgccccccagtcgcccaacccctgcggcgccccggggtgagtcattctccaggtagccagctg
gcgcttgccttttctgtattttatttagacaagagatgggaatgaggtgggaggtggaagaaggagagaaga
aaggtgagtttgagctgccttccctagcttttagacccctgggtgggctctgtgcagtcactggaggtgaa
gccaagtggggtgctgggagggagggagggaggtcactggaaaggggagagcctgctggccaccaccgt
ggaggagggaaggcaaggggggtggaggggtgtggcagtggttttggcaaacgctaagagcccttgcct
ccccatttcccatctgcaccccttctctctcccaaatcaatacactagttgttttacccttggcaaa
aaaaaaaa
```

b)

```
>gi|544063467|NM_001282626|Homo sapiens lamin A/C (LMNA), transcript variant 7, mRNA.
METPSQRRATRSQAQASSTPLSPTRITRLQEKEDLQELNDR LAVYIDRVRSLETENAGLRRLRITESEEVV
SREVSIGIKAAEYELGDARKTLDSVAKERARLQLELSKVREEFKELKARNTKKEGDLIAAQARLKDLEAL
LNSKEAALSTALSEKRTLLEGELHDLRGQVAKLEAALGEAKKQLQDEMLRRVDAENRLQTMKEELDFQKNI
YSEELRETKRRRHETRLVEIDNGKQREFESRLADALQELRAQHEDQVEQYKKELEKTYSAKLONARQSAER
NSNLVGAHEELQQSRIRIDSLSAQLSQLQKLAKEAKLRDLEDSLARERDTSRLLAEKEREMAEMRA
RMQQQLDEYQELLDIKLALDMEIHAYRKLEGEERLRLSPSPTSQSRGRASSHSSQTQGGGSVTKRRK
LESTERSSFSQHARTSGRVAVEEVDEEGKFVRLRNKSNEDQSMGNWQIKRQNGDDPLLTYYRFPFKFTLK
AGQVVTIWAAGAGATHSPPTDLVWKAQNTWGCNLSRLTALINSTGEEVAMRKLVRVSVTVVDEDEDEGDD
LLHHHNGSHCSSGDPAYNLRSTVLCGTCGQPADKASAGSGAQSPQNC SIM
```

Figure 2.4 Fasta format file which download from database. a) DNA fasta file contains GI number, Genbank accession number, definition and DNA sequence. b) amino acid fasta file contains GI number, Genbank accession number, definition and amino acid sequence.

Table 2.1 Table structure of database**a) The structure of ‘cardio_GB’ table**

Name of column	Type of data
Accession	varchar(20)
Gi	varchar(20)
Definition	text
bp_length	text
gc_rate	text
Origin	longtext
AAseq	longtext

b) The structure of ‘cardio_clinVar’ table

Name of column	Type of data
Variation	varchar(100)
Gene	text
Disease	text
Clinical_significance	text
var_type	text

c) The structure of ‘cardio_uniProt’ table

Name of column	Type of data
Entry	varchar(20)
Protein	text
Gene	text
Length	int
Molecular_Weight	int
Structure	text
Domain	text
Function	text
Post_Translational_Modification	text
Subcellular_Location	text

The database consists of three tables, and the information of each table is interlocked. The primary key of ‘cardio_GB’ table is ‘Accession’, ‘cardio_clinVar’ table is ‘Variation’ and ‘cardio_uniProt’ table is ‘Entry’

2.3 차별발현 유전자 분석

마이크로어레이를 통한 실험군과 대조군 사이의 유전자 발현 비율(intervention fold difference)분석으로 두 비교군 사이에 발현 량 차이가 유의한 유전자(differentially expressed genes, DEGs)목록을 도출하고, 발현량에 유의한 차이를 보이는 유전자들의 네트워크를 구성한다. 본 연구에 사용된 마이크로어레이 데이터는 NCBI GEO (gene expression omnibus)의 Series GSE1145 데이터로, title은 “changes in cardiac transcription profiles brought about by heart failure”, organism은 Homo sapiens이고 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array로 실험이 진행 된 결과들이다. 실험군 검체는 다양한 심근증으로 인한 심부전으로 심장이식술 과정 중 환자의 심장에서 채취되었고, 대조군 검체는 정상인의 심장에서 채취 되었다. 본 연구에 사용된 데이터들은 <Table 2.2>를 통해 확인 할 수 있다. 위의 데이터들은 NCBI GEO의 다운로드 페이지(<https://www.ncbi.nlm.nih.gov/geo/download/?acc=GDS651>)에서 확장자가 ‘.cel’인 형식의 파일을 다운로드 받았다.

수집한 유전자 발현량 데이터는 통계분석을 하기 전에 standard Affymetrix algorithm인 MAS5 (Microarray Suite 5.0)로 전처리 하였으며 R을 이용하여 데이터 전처리를 진행하였다. MAS5는 regional adjustment를 통한 background correction, 스케일링기법을 통한 normalization, Tukey biweight average법을 통한 summarization의 순서로 진행되고, 그 후에는 발현강도를 \log_2 스케일로 변환시켜 t-test를 하기에 적합하게 데이터들이 분포시키는 한편 fold change값을 구하기 쉽게 해준다. 전처리가 끝난 뒤에는 flag값을 이용하여 발현이 잘 안된 유전자의 결과를 제외 해 준다. 마이크로어레이 칩을 스캐너가 스캔을 하면 spot의 모양, saturation여부, background등의 기준에 따라

flag는 P (Present), M (Marginal), A (Absent) 이렇게 3개의 level로 나뉘는데, 발현이 잘 안된 유전자를 제외하기 위해 모든 sample에서 flag가 'A'인 유전자를 제외하였다. 이렇게 얻어진 데이터를 가지고 통계 분석을 실시 하였는데, 실험군과 대조군의 비교는 t-test로 진행하였다. Normal 검체를 idiopathic dilated, ischemic, hypertrophic cardiomyopathy 환자의 검체 와 각각 비교하였고, 유의 확률은 p-value대신 FDR (False Discovery Rate)을 사용하였다. T-검정 결과 FDR이 0.05이하, 발현량이 2배 이상 차이가 나는 즉, fold change의 절대값이 1보다 큰(two fold change)유전자만 DEGs로 선별 하였다. 확장성 심근증 및 허혈성 심근증, 그리고 비대성 심근증에서 선별한 DEGs는 각각 순서대로 comparison1, comparison 2, comparison 3으로 지정했다.

위에서 지정한 comparison간에 공통적으로 조절되는 유전자 들을 탐색하여 또 다른 유전자들의 목록을 만든다. 확장성 심근증 및 허혈성 심근증, 그리고 비대성 심근증에서 발현이 증가 혹은 감소하는 유전자들 중, 세 그룹에서 공통으로 발현이 조절되는 유전자를 탐색하여 이를 comparison4로 지정한다. 연구에 사용된 샘플들은 여러 원인으로 인한 심근증 환자들 중 심부전의 예후를 나타낸 환자들에게서 채취 되었으므로 comparison1,2,3 모두에서 공통적으로 발현이 조절되는 유전자들인 comparison4는 다양한 원인과 종류의 심근증에서 심부전의 예후를 나타나게 하는 경로에 속해 있는 유전자들이라 할 수 있을 것이다. <Table 2.3>을 통해 데이터 세트를 확인할 수 있다..

Table 2.2 GEO Series GSE1145 data that used for this study

GEO Accession	disease	title	sample type	source name	release date
GSM18442	normal	PGA_PA-N_112	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18443	normal	PGA_PA-N_118	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18444	normal	PGA_PA-N_148	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18445	normal	PGA_PA-N_200	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18446	normal	PGA_PA-N_249	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18447	normal	PGA_PA-N_291	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18448	normal	PGA_PA-N_294	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18449	normal	PGA_PA-N_300	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18450	normal	PGA_PA-N_322	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18451	normal	PGA_PA-N_325	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18452	normal	PGA_PA-N_326	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18477	ICP	PGA_PA-S_115	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18478	ICP	PGA_PA-S_224	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18479	ICP	PGA_PA-S_233	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18480	ICP	PGA_PA-S_242	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18481	ICP	PGA_PA-S_267	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18482	ICP	PGA_PA-S_281	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18483	ICP	PGA_PA-S_297	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18484	ICP	PGA_PA-S_314	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18485	ICP	PGA_PA-S_342	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18486	ICP	PGA_PA-S_386	RNA	Left Ventricle Biopsy	24-Mar-04
GSM18487	ICP	PGA_PA-S_464	RNA	Left Ventricle Biopsy	24-Mar-04
GSM25801	DCM	PGA-Hs_D-1	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25802	DCM	PGA-Hs_D-2	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25803	DCM	PGA-Hs_D-3	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25804	DCM	PGA-Hs_D-4	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25805	DCM	PGA-Hs_D-5	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25806	DCM	PGA-Hs_D-6	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25807	DCM	PGA-Hs_D-7	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25808	DCM	PGA-Hs_D-9	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25809	DCM	PGA-Hs_D-10	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25810	DCM	PGA-Hs_D-12	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25811	DCM	PGA-Hs_D-13	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25812	DCM	PGA-Hs_D-15	RNA	Left Ventricle Biopsy	03-Dec-04

Table 2.2 GEO Series GSE1145 data that used for this study (cont.)

GEO Accession	disease	title	sample type	source name	release date
GSM25818	HCM	PGA-Hs_H-1	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25819	HCM	PGA-Hs_H-2	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25820	HCM	PGA-Hs_H-3	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25821	HCM	PGA-Hs_H-4	RNA	Left Ventricle Biopsy	03-Dec-04
GSM25822	HCM	PGA-Hs_H-5	RNA	Left Ventricle Biopsy	03-Dec-04

GEO Series GSE1145 contains 107 samples, and 39 samples were used in this study, taking into account the platform and phenotype.

DCM : dilated cardiomyopathy, ICM : ischemic cardiomyopathy,

HCM : hypertrophic cardiomyopathy

Table 2.3 Dataset

	Control group	Experimental group
comparison1	normal (11)	idiopathic dilated cardiomyopathy (12)
comparison2	normal (11)	ischemic cardiomyopathy (11)
comparison3	normal (11)	hypertrophic cardiomyopathy (5)
< Compared comparison >		
comparison4	comparison1, comparison2, comparison3	

Genes with different expression levels in the experimental and control groups are selected to construct comparison1, 2 and 3. The genes that are commonly included in comparison1, 2, 3 are selected to form comparison4.

2.4 유전자 네트워크 분석

먼저 open source bioinformatics 소프트웨어이자 데이터베이스인 GN을 통해 DEGs 목록에 있는 유전자들의 네트워크를 구성한다. The Genotype-Tissue Expression v5 (GTEx v5) project 그룹의 Heart-left Ventricle mRNA 타입의 데이터를 기반으로 comparison4에 속하는 유전자들 사이의 발현 상관성으로 이루어진 correlation matrix를 얻는다. GTEx project는 유전자 발현이 인체의 여러 기관에서 어떻게 조절되는지를 체계적으로 연구하기 위해 발현 상관성 데이터를 제공한다. Version5까지는 175개체로부터 43개의 조직을 통해 1641개의 RNA sequencing 데이터를 얻었으며, 연구에 사용된 Heart-left Ventricle mRNA type의 데이터는 246개의 샘플을 통해 만들어 졌다 <Table 2.4>(GTEx Consortium, 2015). 이 matrix를 이용하여 유전자를 노드로, 유전자간 interaction을 연결선으로 하는 네트워크를 구성하는데, 유전자간 발현 정도의 상관성에서 피어슨 상관계수의 절대값이 0.5 및 0.6이상인 interaction만 연결선을 이루도록 한다. 네트워크의 구성 및 분석은 R의 sna 패키지를 이용하여 수행하였다.

처음으로 네트워크의 전체적인 형태를 살펴보기 위해 R의 sna 패키지를 이용하여 피어슨 상관계수의 절대값이 0.7이상인 interaction만 연결선을 이루도록 하여 가중치가 없는 무방향 네트워크를 구성하였다. 가중치가 없다는 것은 노드a와 노드b 사이의 interaction의 정도가 일정 수치 이상이면 연결선을 형성하는 것을 말하며, 무방향 네트워크는 노드a에서 노드b로 가는 연결선이 있으면 노드b에서 노드a로 가는 연결선도 함께 형성되는 네트워크를 말한다. 다음으로는 네트워크 분석을 위해 피어슨 상관계수의 절대값이 0.5이상인 interaction만 연결선을 이루도록 하여 가중치가 없는 무방향 네트워크를 구성한다. 마지막으로 comparison4로 구성된 네트워크

와 확장성 및 비대성 심근증 환자에서 선천적인 변이를 보이는 유전자들과의 관계를 알아보기 위해 2부 네트워크를 구성하였다. 2부 네트워크는 서로 성질이 다른 개체간의 관계로 구성된 네트워크로 사회 구성원과 조직간 관계를 나타내는 멤버십 자료가 대표적인데 예를 들어 사람 2명이 어느 조직에 속해있으면 서로 연결된 것으로 간주하고, 두 기관이 한 명의 공통으로 소속된 사람이 있으면 서로 연결된 것으로 간주하는 개념이다(Roberts, 2000). 본 연구에서는 comparison4로 구성된 네트워크를 조직, 네트워크와 확장성 및 비대성 심근증 환자에서 선천적인 변이를 보이는 유전자들을 사회 구성원으로 비유하여 2부 네트워크를 구성했고, 해당 네트워크의 행렬은 다음과 같다.

$$B = \begin{pmatrix} 0_n & A \\ A^t & 0_m \end{pmatrix}$$

A : 선천적인 변이를 보이는 유전자들을 행, comparison4의 유전자들을 열로 지정하여 구성한 행렬

B : 전체 네트워크 행렬((n+m)*(n+m))

n : 행 노드, 즉 선천적인 변이를 보이는 유전자들의 개수

m : 열노드, 즉 comparison4의 유전자들의 개수

Comparison4의 유전자들과 확장성 및 비대성 심근증 환자에서 선천적인 변이를 보이는 유전자들에 대한 2부 네트워크를 각각 구성 하였으며, 2부 네트워크를 구성하는데 사용된 확장성 및 비대성 심근증 환자에서 선천적인 변이를 보이는 유전자들은 <Table 2.5>에서 확인 할 수 있다. 본 연구를 위해 구성한 네트워크들의 목록은 <Table 2.6>에서 확인 할 수 있다.

다음은 위의 방법으로 구성된 네트워크를 R의 sna 패키지를 이용하여 네트워크의 밀도 및 중심성 분석을 수행하였다. 먼저 comparison4의 유전자로 구성된 네트워크가 얼마나 유효하고 효과적인

인지 알아보기 위해 comparison4의 가중치가 없는 무방향 네트워크인 network2로 네트워크의 밀도를 구한다. 네트워크의 밀도는 가능한 연결선의 수 대비 총 연결선의 수로 정의가 되며 구하는 식은 아래와 같다.

$$\text{network density} = \frac{i}{{}_nC_2}$$

i: 실제 총 연결수

$${}_nC_2 : n*(n-1)/2$$

n: 노드의 수

중심성 분석은 연결정도 중심성, 근접중심성, 중개중심성 이렇게 세가지로 이루어져 있으며 네트워크의 밀도와 마찬가지로 R의 sna 패키지를 이용하여 분석하였다. 연결정도 중심성은 각 노드가 가지고 있는 연결선의 수로 측정하였다. 한 노드가 다른 노드들과 평균적으로 얼마나 가깝게 위치하는지 알 수 있는 지표인 근접중심성은 다음의 식으로 구할 수 있다.

$$C(i) = \frac{1}{\sum_{j \neq i} d(i, j)}$$

C(i): 근접중심성

d(i, j) : 두 노드 사이의 거리

n: 노드의 수

하지만 어느 한 노드가 다른 노드로부터 도달되어지는 연결선이 없다면 근접중심성은 0이 되므로 이를 보완하기 위해 조화중심성이 사용된다. 조화중심성은 아래와 같이 구하며, 본 연구에서 구성한 네트워크에는 모든 노드들과 연결선을 형성하지 않는 노드들이 존재 하기 때문에 근접중심성 대신 조화중심성을 사용하였으며, 이후에 나오는 근접중심성은 모두 조화중심성을 뜻한다(Dekker, 2005).

$$C(h) = \sum_{j \neq i} \frac{1}{d(i, j)}$$

$C(h)$: 조화중심성

$d(i, j)$: 두 노드 사이의 거리

n : 노드의 수

한 노드가 두 개의 다른 노드의 가장 짧은 경로에 위치하여 허브로서의 역할을 하는지를 정량적으로 나타내 주는 중개중심성은 다음의 식으로 구할 수 있다.

$$C(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$C(v)$: 중개중심성

σ_{st} : 노드 s 에서 노드 t 까지 최단경로의 수

$\sigma_{st}(v)$: 노드 s 에서 노드 t 까지 최단경로 중 노드 v 를 지나는 경로의 수

이렇게 구한 중심성의 지표들은 네트워크의 규모에 따라 차이가 나기 때문에 수치를 정규화 해 주어야 한다. 연결정도 중심성과 근접중심성은 자기자신을 제외한 모든 노드들이 연결되어 있을 때의 값($n-1$, n =전체 노드의 수)으로 나누어 주고, 중개중심성은 자기자신을 제외하고 짝을 이룰 수 있는 값으로 나누어 주는데 n 을 전체 노드의 수라고 한다면, 방향성이 있는 네트워크에서는 $(n-1)(n-2)$ 로 나누고 방향성이 없는 네트워크에서는 $(n-1)(n-2)/2$ 로 나눈다.

위의 중심성 분석은 network2,3,4로 수행 하였는데 2부 네트워크인 network3,4에서는 comparison4의 유전자들에 대해 선천적으로 심근증과 관련이 있는 유전자들과 comparison4의 유전자들을 기준으로 각각 중심성 분석을 수행하였다.

Table 2.4 Sample list of left ventricle in GTEv5

Tissue sample Id	Sex	Age Bracket	Hardy Scale
GTEX-111FC-0826	male	60-69	Fast death - violent
GTEX-111YS-0426	male	60-69	Ventilator case
GTEX-1122O-0826	female	60-69	Ventilator case
GTEX-117YW-0326	male	50-59	Intermediate death
GTEX-117YX-1126	male	50-59	Ventilator case
GTEX-11DXX-0326	female	60-69	Ventilator case
GTEX-11DXY-0826	male	60-69	Fast death - natural causes
GTEX-11DXZ-0626	male	50-59	Ventilator case
GTEX-11EM3-0626	female	20-29	Ventilator case
GTEX-11EMC-0726	female	60-69	Fast death - natural causes
GTEX-11GS4-0526	male	60-69	Fast death - natural causes
GTEX-11GSP-1326	female	60-69	Fast death - natural causes
GTEX-11I78-0826	female	50-59	Ventilator case
GTEX-11LCK-0926	male	30-39	Ventilator case
GTEX-11O72-1126	male	40-49	Fast death - natural causes
GTEX-11ONC-0126	male	60-69	Fast death - natural causes
GTEX-11TT1-1326	male	20-29	Ventilator case
GTEX-11TUW-1026	male	60-69	Fast death - natural causes
GTEX-11ZUS-0226	male	30-39	Fast death - violent
GTEX-11ZVC-0426	female	50-59	Fast death - natural causes
GTEX-1211K-0626	female	60-69	Ventilator case
GTEX-1212Z-0626	male	60-69	Fast death - natural causes
GTEX-12584-0926	male	40-49	Intermediate death
GTEX-12696-1126	male	60-69	Intermediate death
GTEX-1269C-0826	female	60-69	Fast death - natural causes
GTEX-12BJ1-0226	male	60-69	Ventilator case
GTEX-12WSC-1126	male	50-59	Fast death - natural causes
GTEX-12WSD-1126	female	60-69	Fast death - natural causes
GTEX-12WSG-0526	female	50-59	Ventilator case
GTEX-12WSK-0626	female	40-49	Ventilator case
GTEX-12ZZW-1026	male	50-59	Slow death
GTEX-12ZZY-1126	male	60-69	Fast death - natural causes
GTEX-12ZZZ-1726	male	60-69	Fast death - natural causes
GTEX-13112-0426	male	50-59	Fast death - natural causes
GTEX-1313W-1426	female	50-59	Fast death - natural causes
GTEX-1314G-0126	male	30-39	Ventilator case
GTEX-131XE-0626	male	50-59	Ventilator case
GTEX-131XF-0326	male	60-69	Ventilator case
GTEX-131XG-0626	female	50-59	Ventilator case
GTEX-131XH-0226	male	50-59	Fast death - natural causes
GTEX-132AR-2026	female	50-59	Ventilator case
GTEX-132NY-1726	male	60-69	Fast death - violent
GTEX-132QS-0426	male	60-69	Ventilator case

Table 2.4 Sample list of left ventricle in GTEx v5 (cont.)

Tissue sample Id	Sex	Age Bracket	Hardy Scale
GTEX-1339X-0426	male	40-49	Ventilator case
GTEX-1399R-1926	male	30-39	Ventilator case
GTEX-1399S-1026	female	30-39	Ventilator case
GTEX-1399T-0426	male	60-69	Fast death - natural causes
GTEX-1399U-0526	female	50-59	Ventilator case
GTEX-139UC-0426	male	60-69	Fast death - natural causes
GTEX-139YR-0826	male	50-59	Ventilator case
GTEX-13CF2-0426	male	60-69	Fast death - natural causes
GTEX-13CF3-0726	female	60-69	Ventilator case
GTEX-13D11-1726	female	50-59	Ventilator case
GTEX-13FH7-0926	female	50-59	Ventilator case
GTEX-13FHP-1026	male	50-59	Slow death
GTEX-13FLV-0926	male	50-59	Fast death - natural causes
GTEX-13FTW-1026	male	40-49	Ventilator case
GTEX-13FTZ-0226	male	60-69	Fast death - natural causes
GTEX-13JVG-1326	male	60-69	Fast death - natural causes
GTEX-13N11-0426	female	50-59	Ventilator case
GTEX-13N1W-1026	male	70-79	Ventilator case
GTEX-13NYB-0226	male	40-49	Fast death - natural causes
GTEX-13NZB-0226	male	30-39	Ventilator case
GTEX-13O1R-1226	male	60-69	Fast death - natural causes
GTEX-13O21-2326	male	50-59	Ventilator case
GTEX-13O3O-1526	female	60-69	Intermediate death
GTEX-13O3P-1626	male	50-59	Slow death
GTEX-13O61-0526	male	60-69	Ventilator case
GTEX-13OVG-0526	male	60-69	Ventilator case
GTEX-13OVL-0526	male	50-59	Fast death - natural causes
GTEX-13OW5-1126	male	60-69	Intermediate death
GTEX-13OW6-1126	male	50-59	Fast death - natural causes
GTEX-13OW8-2126	male	60-69	Fast death - natural causes
GTEX-13PVQ-1226	male	50-59	Fast death - natural causes
GTEX-13PVR-1026	female	60-69	Ventilator case
GTEX-13QBU-0426	female	40-49	Ventilator case
GTEX-13RTJ-0726	male	60-69	Slow death
GTEX-13S86-0226	male	40-49	Ventilator case
GTEX-13U4I-0826	female	40-49	Ventilator case
GTEX-13VXT-1126	female	20-29	Ventilator case
GTEX-13W3W-0426	female	60-69	Ventilator case
GTEX-13X6H-0426	female	50-59	Ventilator case
GTEX-13X6I-1426	male	60-69	Slow death
GTEX-13X6K-1826	female	60-69	Slow death
GTEX-13YAN-1326	male	60-69	Ventilator case
GTEX-145LV-0326	male	50-59	Ventilator case

Table 2.4 Sample list of left ventricle in GTEx v5 (cont.)

Tissue sample Id	Sex	Age Bracket	Hardy Scale
GTEX-145MN-0626	male	30-39	Ventilator case
GTEX-145MO-1126	male	60-69	Intermediate death
GTEX-146FH-1026	female	50-59	Ventilator case
GTEX-146FR-0626	female	60-69	Ventilator case
GTEX-147F4-1226	male	50-59	Fast death - natural causes
GTEX-148VI-0326	female	60-69	Ventilator case
GTEX-148VJ-1526	male	70-79	Fast death - natural causes
GTEX-1497J-1226	male	60-69	Ventilator case
GTEX-14A5I-1226	female	50-59	Slow death
GTEX-14A6H-0226	male	30-39	Ventilator case
GTEX-14C38-1426	male	50-59	Ventilator case
GTEX-14C5O-1326	male	60-69	Slow death
GTEX-14E6D-0926	male	50-59	Fast death - natural causes
GTEX-14E7W-1226	male	70-79	Fast death - natural causes
GTEX-N7MS-0826	male	60-69	Fast death - natural causes
GTEX-N7MT-1326	female	60-69	Slow death
GTEX-NFK9-0926	male	40-49	Ventilator case
GTEX-NPJ8-0426	male	40-49	Slow death
GTEX-O5YT-0326	male	20-29	Ventilator case
GTEX-O5YV-0326	female	60-69	Intermediate death
GTEX-O5YW-0326	male	50-59	Ventilator case
GTEX-OHPJ-0926	male	60-69	Ventilator case
GTEX-OHPK-0326	female	50-59	Ventilator case
GTEX-OHPL-0326	female	60-69	Ventilator case
GTEX-OHPM-0326	male	50-59	Ventilator case
GTEX-OIZF-0326	male	60-69	Ventilator case
GTEX-OIZH-0326	male	50-59	Ventilator case
GTEX-OOBJ-0326	male	60-69	Ventilator case
GTEX-OOBK-3025	male	40-49	Ventilator case
GTEX-OXRK-0826	female	50-59	Ventilator case
GTEX-OXRL-0326	male	50-59	Ventilator case
GTEX-OXRO-2026	female	60-69	Slow death
GTEX-P44G-0826	male	30-39	Ventilator case
GTEX-P44H-0726	male	40-49	Slow death
GTEX-P4PP-0326	female	30-39	Ventilator case
GTEX-P4PQ-0326	male	60-69	Ventilator case
GTEX-P4QS-0326	male	60-69	Ventilator case
GTEX-P78B-0426	female	40-49	None
GTEX-PLZ5-0626	male	50-59	Ventilator case
GTEX-POMQ-0326	female	20-29	Ventilator case
GTEX-PSDG-0926	male	50-59	Ventilator case
GTEX-PVOW-0426	male	40-49	Fast death - natural causes
GTEX-PWCY-0526	female	20-29	Ventilator case

Table 2.4 Sample list of left ventricle in GTEx v5 (cont.)

Tissue sample Id	Sex	Age Bracket	Hardy Scale
GTEX-PWOO-0526	female	40-49	Ventilator case
GTEX-PX3G-0326	female	20-29	Ventilator case
GTEX-Q2AH-0526	male	40-49	Ventilator case
GTEX-QDVJ-0426	male	50-59	Ventilator case
GTEX-QDVN-0326	male	50-59	Ventilator case
GTEX-QEG4-0426	male	30-39	Fast death - violent
GTEX-QEG5-0926	male	20-29	Ventilator case
GTEX-QEL4-0926	male	50-59	Intermediate death
GTEX-QESD-0526	male	20-29	Ventilator case
GTEX-QLQ7-0526	male	60-69	Ventilator case
GTEX-QMRM-0526	male	50-59	Ventilator case
GTEX-QV44-0526	male	50-59	Ventilator case
GTEX-QVJO-1926	female	60-69	Slow death
GTEX-QXCU-0826	male	50-59	Fast death - violent
GTEX-R45C-0926	female	60-69	Ventilator case
GTEX-R53T-0926	female	50-59	Ventilator case
GTEX-R55C-0326	male	40-49	Ventilator case
GTEX-R55E-1026	male	20-29	Fast death - violent
GTEX-R55G-0526	female	40-49	Ventilator case
GTEX-REY6-1026	male	60-69	Ventilator case
GTEX-RN64-0826	male	50-59	Fast death - violent
GTEX-RNOR-0826	female	50-59	Fast death - violent
GTEX-RTLS-0826	female	60-69	Slow death
GTEX-RU72-0326	female	50-59	Fast death - natural causes
GTEX-RUSQ-0526	male	50-59	Ventilator case
GTEX-RWS6-0326	female	60-69	Ventilator case
GTEX-RWSA-0626	male	40-49	Fast death - violent
GTEX-S32W-0626	female	50-59	Ventilator case
GTEX-S33H-0526	male	30-39	Fast death - natural causes
GTEX-S3XE-0426	male	50-59	Ventilator case
GTEX-S7SF-0526	female	20-29	Ventilator case
GTEX-SE5C-0626	female	40-49	Ventilator case
GTEX-SIU7-0426	male	50-59	Ventilator case
GTEX-SIU8-0826	male	50-59	Fast death - natural causes
GTEX-SNMC-0126	male	20-29	Ventilator case
GTEX-SUCS-0326	male	70-79	Ventilator case
GTEX-T2IS-0426	female	20-29	Fast death - violent
GTEX-T6MN-0926	male	50-59	Fast death - natural causes
GTEX-U3ZH-0326	male	30-39	Ventilator case
GTEX-U3ZN-1426	female	30-39	Ventilator case
GTEX-U4B1-0326	male	40-49	Ventilator case
GTEX-U8XE-1126	male	30-39	Slow death
GTEX-UJHI-0426	female	50-59	Ventilator case

Table 2.4 Sample list of left ventricle in GTEx v5 (cont.)

Tissue sample Id	Sex	Age Bracket	Hardy Scale
GTEX-UJMC-0526	male	50-59	Ventilator case
GTEX-UPK5-0326	male	40-49	Ventilator case
GTEX-V1D1-0526	male	50-59	Ventilator case
GTEX-V955-0726	male	60-69	Ventilator case
GTEX-WEY5-0426	female	40-49	Ventilator case
GTEX-WFG7-0726	male	20-29	Ventilator case
GTEX-WFG8-0626	male	20-29	Ventilator case
GTEX-WFON-0326	male	40-49	Ventilator case
GTEX-WH7G-0426	male	40-49	Ventilator case
GTEX-WHPG-0826	male	50-59	Ventilator case
GTEX-WHSE-0926	male	20-29	Intermediate death
GTEX-WHWD-0426	male	50-59	Ventilator case
GTEX-WI4N-0626	female	40-49	Ventilator case
GTEX-WL46-0926	male	50-59	Fast death - natural causes
GTEX-WQUQ-1426	male	60-69	Ventilator case
GTEX-WRHU-1226	female	50-59	Fast death - violent
GTEX-WWYW-1326	female	50-59	Intermediate death
GTEX-WY7C-0526	male	50-59	Ventilator case
GTEX-WYJK-1026	male	50-59	Ventilator case
GTEX-WZTO-1326	male	40-49	Fast death - violent
GTEX-X15G-0426	female	50-59	Ventilator case
GTEX-X261-1426	male	50-59	Slow death
GTEX-X3Y1-0426	male	40-49	Ventilator case
GTEX-X5EB-0826	male	40-49	Ventilator case
GTEX-X8HC-1626	female	50-59	Ventilator case
GTEX-XBEC-1326	male	50-59	Fast death - natural causes
GTEX-XBED-0526	male	60-69	Ventilator case
GTEX-XGQ4-0326	male	50-59	Ventilator case
GTEX-XPT6-0126	male	20-29	Ventilator case
GTEX-XPVG-0826	male	50-59	Ventilator case
GTEX-XQ3S-0626	male	20-29	Fast death - violent
GTEX-XQ8I-0126	male	50-59	Ventilator case
GTEX-XUJ4-0526	female	60-69	Ventilator case
GTEX-XV7Q-0826	female	40-49	Ventilator case
GTEX-XXEK-0926	male	50-59	Ventilator case
GTEX-Y114-0226	female	50-59	Ventilator case
GTEX-Y3I4-0626	male	50-59	Ventilator case
GTEX-Y3IK-0826	female	50-59	Ventilator case
GTEX-Y5V5-0626	female	60-69	Ventilator case
GTEX-Y5V6-0826	male	60-69	Ventilator case
GTEX-Y8E4-0326	male	40-49	Ventilator case
GTEX-YB5K-0126	female	40-49	Ventilator case
GTEX-YEC3-0726	male	50-59	Ventilator case

Table 2.4 Sample list of left ventricle in GTEx v5 (cont.)

Tissue sample Id	Sex	Age Bracket	Hardy Scale
GTEX-YEC4-0326	male	40-49	Ventilator case
GTEX-YF70-0426	male	50-59	Ventilator case
GTEX-ZAB4-1026	male	40-49	Slow death
GTEX-ZAB5-0226	male	50-59	Ventilator case
GTEX-ZAJG-0826	female	50-59	Intermediate death
GTEX-ZDTS-1326	male	60-69	Fast death - natural causes
GTEX-ZDTT-0726	male	60-69	Ventilator case
GTEX-ZDXO-0726	male	60-69	Fast death - natural causes
GTEX-ZDYS-0226	male	50-59	Ventilator case
GTEX-ZEX8-0426	male	50-59	Ventilator case
GTEX-ZF29-0226	female	60-69	Ventilator case
GTEX-ZF2S-1026	female	50-59	Ventilator case
GTEX-ZF3C-1226	female	50-59	Ventilator case
GTEX-ZG7Y-0926	male	50-59	Ventilator case
GTEX-ZGAY-0326	female	40-49	Ventilator case
GTEX-ZLFU-0226	male	40-49	Ventilator case
GTEX-ZPCL-0426	female	60-69	Ventilator case
GTEX-ZPU1-0626	male	40-49	Ventilator case
GTEX-ZQUD-1526	male	30-39	Ventilator case
GTEX-ZTPG-1226	female	20-29	Ventilator case
GTEX-ZTTD-0526	male	50-59	Intermediate death
GTEX-ZUA1-0826	male	40-49	Fast death - natural causes
GTEX-ZV7C-0326	male	50-59	Ventilator case
GTEX-ZVT2-1026	female	50-59	Ventilator case
GTEX-ZVZP-0226	male	50-59	Ventilator case
GTEX-ZYFC-1126	male	50-59	Fast death - natural causes
GTEX-ZYFG-0426	female	60-69	Ventilator case
GTEX-ZYT6-1726	male	30-39	Fast death - violent
GTEX-ZYW4-0926	male	60-69	Intermediate death
GTEX-ZYWO-0626	female	40-49	Ventilator case
GTEX-ZZPU-0926	female	50-59	Ventilator case

Hardy Scale : Death Circumstances. Death Classification: 4-point Hardy Scale.

- 1) Violent and fast death Deaths due to accident, blunt force trauma or suicide, terminal phase estimated at < 10 min.
- 2) Fast death of natural causes Sudden unexpected deaths of people who had been reasonably healthy, after a terminal phase estimated at < 1 hr (with sudden death from a myocardial infarction as a model cause of death for this category).
- 3) Intermediate death Death after a terminal phase of 1 to 24 hrs (not classifiable as 2 or 4); patients who were ill but death was unexpected.
- 4) Slow death Death after a long illness, with a terminal phase longer than 1 day (commonly cancer or chronic pulmonary disease); deaths that are not unexpected.
- 5) Ventilator Case All cases on a ventilator immediately before death.

Table 2.5 Genes that have germline variant in cardiomyopathy

Subtype of cardiomyopathy	Genes
dilated cardiomyopathy	ABCC9, ACTC1, ACTN2, ANKRD1, BAG3, CAV3, CRYAB, CSRP3, DES, DMD, DSG2, DSP, EMD, EYA4, FHL2, FKRP, FKTN, GATAD1, ILK, LAMA4, LAMP2, LDB3, LMNA, MYBPC3, MYH7, MYPN, NEXN, PDLIM3, PKD1, PKP2, PLN, PRDM16, PSEN1, PSEN2, RBM20, RYR2, SCN5A, SGCD, TAZ, TBX20, TCAP, TMPO, TNNC1, TNNI3, TNNT2, TPM1, TTN, TXNRD2, VCL
hypertrophic cardiomyopathy	ACTC1, ACTN2, ANKRD1, CALR3, CAV3, CSRP3, FHL1, GLA, JPH2, LAMP2, LDB3, MYBPC3, MYH7, MYL2, MYL3, MYLK2, MYOZ2, MYPN, NEXN, PLN, PRKAG2, TCAP, TNNC1, TNNI3, TNNT2, TPM1, TTN, VCL

— Representation of genes involved in dilated cardiomyopathy and hypertrophic cardiomyopathy (Van Tintelen et al., 2014). The above genes were used to construct 2-mode network comparison 4.

Table 2.6 List of network for this study

Network	Explain
network1	Non-weighted and non-directional network that composed of comparison4 genes (Absolute value of Pearson's correlation coefficient is above 0.7)
network2	Non-weighted and non-directional network that composed of comparison4 genes (Absolute value of Pearson's correlation coefficient is above 0.5)
network3	Two-mode network between comparison4 genes and germline variant of dilated cardiomyopathy (Absolute value of Pearson's correlation coefficient is above 0.6)
network4	Two-mode network between comparison4 genes and germline variant of hypertrophic cardiomyopathy (Absolute value of Pearson's correlation coefficient is above 0.6)

— The node is the gene and the connecting line is absolute value of Pearson's correlation coefficient in the network.

2.5 단백질 네트워크 분석

Comparison4와 확장성 및 비대성 심근증의 원인이 되는 유전자들 간에 존재하는 protein interaction을 알아보기 위해 Search Tool for the Retrieval of Interacting Genes/Proteins (<http://string-db.org/>, STRING)의 combined score를 바탕으로 2부 네트워크를 구성하여 중심성 분석을 실시한다. STRING은 단백질들간의 상호작용을 예측해 주는 웹 기반의 생물학적 데이터베이스로, 실험 데이터와 계산 예측법 및 텍스트 수집 등 다양한 출처의 정보들을 수록하고 있다(Szklarczyk et al., 2015). ‘Experimental/Biochemical Data’, ‘Co-Expression’, ‘Gene Fusions’ 등 기능적 연관성을 나타내주는 지표들을 통합한 Combined Score가 0.4이상인 interaction을 기반으로 comparison4와 확장성 및 비대성 심근증의 원인 유전자들 사이에 2부 네트워크를 구성하고, comparison4와 확장성 심근증 사이의 2부네트워크는 network5로, comparison4와 비대성 심근증 사이의 2부네트워크는 network6으로 각각 명명한다. 이렇게 구성한 단백질 네트워크는 R 프로그램의 ‘sna’package를 이용하여 연결정도 중심성, 근접중심성, 중개중심성을 척도로 하는 중심성 분석을 한 뒤, 유전자 네트워크 분석 결과와 통합하기 위해 각 중심성 척도에서 각각의 유전자가 차지하는 비율로 유전자의 중심성을 나타낸다.

제 3 장. 연구 결과

3.1 심근증에 대한 유전학적 정보 검색 시스템 구축

본 연구에서는 GenBank (www.ncbi.nlm.nih.gov/genbank/), Clin Var (www.ncbi.nlm.nih.gov/clinvar/), UniProt (www.uniprot.org/)에서 심근증에 대한 정보를 수집하여 심근증과 관련이 있는 유전자들의 염기서열 및 아미노산서열, 해당 유전자의 변이와 이 변이의 임상적 의의, 유전자가 발현하는 단백질과 단백질의 구조에 대한 정보들을 통합한 데이터베이스의 구축을 시도 하였다. 위의 데이터베이스를 기반으로 한 심근증의 유전학적 정보 검색 시스템은 CGDB (Cardiomyopathy Genetic Database)라고 칭하였고 서열정보 782개, 유전자 변이에 대한 정보 3919개, 단백질에 대한 정보 242개를 수록하고 있다. <Figure 3.1>은 구축한 검색 시스템의 메인 페이지로 데이터베이스 구축 목적과 수록되어 있는 데이터 정보를 명시하였다. CGDB의 주요 기능은 첫 번째, 심근증에 대한 기본적인 의학적 정보를 얻을 수 있는데, 이는 미국 NIH (National Institutes of Health)에서 제공하는 자료이며 <Figure 3.2>를 통해 확인 할 수 있다. 두 번째, 유전자의 염기서열, 유전 변이, 발현 단백질 등의 유전학적 정보를 제공하며 마지막으로 염기 및 아미노산 서열을 FASTA 형식의 파일로 다운로드 할 수 있는 시스템을 제공한다.

심근증의 유전학적 정보를 얻는 방법은 크게 두 가지로 나뉘는데, 첫 번째는 <Figure 3.3>같이 심근증의 아형 별로 임상적 의의

가 ‘Pathogenic’ 이나 ‘Likely pathogenic’인 유전 변이를 갖는 유전자들의 목록에서 선택하는 것이다. 유전자를 선택하면 <Figure 3.4>와 같이 해당 아형과 관련이 있는 유전자의 변이 목록을 얻을 수 있으며 어떠한 변이가 얼마나 있는지 알아 볼 수 있다. 또한 <Figure 3.5>와 같이 해당 유전자가 발현하는 단백질의 구조에 대한 정보도 얻을 수 있는데 이를 통해 유전변이가 단백질의 어떠한 부분에 영향을 미쳐서 질병의 원인이 되는지 알아 볼 수 있다. 유전학적 정보를 얻는 두 번째 방법은 <Figure 3.6>과 같이 알아보하고자 하는 변이의 조건을 선택하여 이에 해당하는 정보를 얻는 방법이 있다. 각각의 아형과 관련이 있는 유전자, 임상적 의의, 변이의 종류를 선택하면 이에 해당하는 유전학적 정보들을 얻을 수 있다. 이 기능을 이용하면 <Figure 3.7>과 같이 특정 조건에 부합하는 유전자들의 목록을 구할 수 있다. 특히 ‘Gene symbol’을 ‘All genes’로 선택하고 ‘Clinical significance’를 ‘Pathogenic’ 이나 ‘Likely pathogenic’, Variant Type을 ‘All types’로 선택한 뒤 검색을 하면 각 아형 별로 임상적으로 의의가 있는 유전자들의 목록을 얻을 수 있는데, 이는 여러 유전자의 변이 여부를 한번에 검사 할 수 있는 NGS를 통한 유전자 선별검사에 필요한 유전자 패널을 구성하는데 이용 할 수 있다. 이렇게 얻은 유전자 목록은 <Table 3.1>에서 확인 할 수 있다.

마지막으로 염기 및 아미노산 서열을 FASTA 형식의 파일로 다운로드 할 수 있는데 <Figure 3.8>의 페이지에서 각 아형에 해당하는 유전자 목록 중 원하는 유전자를 선택하면 <Figure 3.9>에서처럼 선택한 유전자에 해당하는 서열 정보를 얻을 수 있으며 각각의 서열정보는 FASTA형식의 파일로 다운로드 할 수 있다.

Cardiomyopathy Genetic Database


Home	Cardiomyopathy?	Genes	Search	Download
				
<p>Welcome To CGDB in LCBB</p> <ul style="list-style-type: none"> - CGDB(Cardiomyopathy Genetic Database) contains information from clinical research and other types of data on variants in genes related to Cardiomyopathy. - This data are provided to enhance knowledge and encourage progress in the scientific community and are to be used only for research and educational purposes. - Mutation nomenclature follows the standard reported here : http://www.hgvs.org/mutnomen/ - Numbering of nucleotides starts at the first position of the methionine start-codon. 				
<p><small>Copyright 2017 Lab of Computational Biology and Bioinformatics All rights reserved Seoul National University, Korea +82-2-850-2552</small></p>				

Figure 3.1 Front page of database. CGDB is a database for genetic information of cardiomyopathy. The main menu consists of ‘Home’, ‘Cardiomyopathy?’, ‘Genes’, ‘Search’ and ‘Download’.

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
<p>Explore Cardiomyopathy</p> <div> <div> <p>WHAT IS...</p> <p>TYPES</p> <p>OTHER NAMES</p> <p>CAUSES</p> <p>WHO IS AT RISK</p> <p>SIGNS / SYMPTOMS</p> <p>DIAGNOSIS</p> <p>TREATMENTS</p> <p>PREVENTION</p> </div> <div> <h3>What Is Cardiomyopathy?</h3> <p>Cardiomyopathy refers to diseases of the heart muscle. These diseases have many causes, signs and symptoms, and treatments. In cardiomyopathy, the heart muscle becomes enlarged, thick, or rigid. In rare cases, the muscle tissue in the heart is replaced with scar tissue.</p> <p>As cardiomyopathy worsens, the heart becomes weaker. It's less able to pump blood through the body and maintain a normal electrical rhythm. This can lead to heart failure or irregular heartbeats called arrhythmias. In turn, heart failure can cause fluid to build up in the lungs, ankles, feet, legs, or abdomen.</p> <p>The weakening of the heart also can cause other complications, such as heart valve problems.</p> <h4>Overview</h4> <p>The types of cardiomyopathy are:</p> <ul style="list-style-type: none"> - Hypertrophic cardiomyopathy - Dilated cardiomyopathy - Restrictive cardiomyopathy - Arrhythmogenic right ventricular dysplasia cardiomyopathy - Left ventricular noncompaction cardiomyopathy - Unclassified cardiomyopathy <p>Cardiomyopathy can be acquired or inherited. "Acquired" means you aren't born with the disease, but you develop it due to another disease, condition, or factor. "Inherited" means your parents passed the gene for the disease on to you. Many times, the cause of cardiomyopathy isn't known.</p> <p>Cardiomyopathy can affect people of all ages. However, people in certain age groups are more likely to have certain types of cardiomyopathy. This article focuses on cardiomyopathy in adults.</p> <h4>Outlook</h4> <p>Some people who have cardiomyopathy have no signs or symptoms and need no treatment. For other people, the disease develops quickly, symptoms are severe, and serious complications occur.</p> <p>Treatments for cardiomyopathy include lifestyle changes, medicines, surgery, implanted devices to correct arrhythmias, and a non-surgical procedure. These treatments can control symptoms, reduce complications, and slow the disease from getting worse.</p> </div> </div>				
<p><small>Copyright 2017 Lab of Computational Biology and Bioinformatics All rights reserved Seoul National University, Korea +82-2-850-2552</small></p>				

Figure 3.2 Medical information page. It contains medical information that provided by U.S. NIH (National Institutes of Health).

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
------	-----------------	-------	--------	----------

Genes related to cardiomyopathy

- Include genes which clinical significance is "Pathogenic" or "Likely pathogenic"
- Click the gene name to investigate variants

< Dilated cardiomyopathy >												
ABCC9	ACTC1	ACTN2	BAG3	CRYAB	DES	DMD	DNAJB5	DSG2	DSB	EYA4	FKTN	GATAO1
LAMA4	LOB3	LMNA	MYBPC3	MYH6	MYH7	MYPN	NEXN	PLN	PRDM16	PSEN1	RAF1	RBM2
SCN5A	SCD2	SDHA	SGCD	TCAP	TNNC1	TNNI3	TNNI2	TPM1	TSPM	TTN	VCL	
< Hypertrophic cardiomyopathy >												
ACTC1	ACTN2	CAV3	CBSP3	DES	PHL1	JPH2	LAMP2	MYBPC3	MYH6	MYH7	MYL2	MYL3
MYLK2	MYO6	MYOZ2	NEXN	PLN	PMPCA	PRKAG2	RAF1	TNNC1	TNNI3	TNNI2	TPM1	TTN
VCL												
< Restrictive cardiomyopathy >												
MYH7	MYPN	TNNI3	TNNI2									
< Arrhythmogenic right ventricular cardiomyopathy >												
DSG2	DSG2	DSB	JUP	PKP2	RYR2	TGFB3	TMEM43					
< Left ventricular noncompaction cardiomyopathy >												
MYH7	RYR2	SCN5A	TTN									

Copyright 2017 Lab of Computational Biology and Bioinformatics All rights reserved
Seoul National University, Korea +82-2-550-2552

Figure 3.3 List for subtype of cardiomyopathy. It include genes which clinical significance is 'Pathogenic' or 'Likely pathogenic'.

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
------	-----------------	-------	--------	----------

Variant of *TTN* related to cardiomyopathy

Total variant : 625
 - deletion : 98
 - duplication : 36
 - indel : 3
 - insertion : 7
 - single nucleotide variant : 410 (nonsense : 105, a.a. change : 285, synonymous : 7)
 - splice site : 38
 - frameshift : 131
 - nonsense : 116
 - etc : 33

Variation	Gene	Disease	Clinical_Significance	Variation_Type	Protein
NIL001286880.1(TTN):C.114540G>T (P.A5138(S>P))	TTN	DCM	Conflicting Interpretations	deletion frameshift	CRKX
NIL001286880.1(TTN):C.12187C>T (P.G140237R)	TTN	DCM	Pathogenic	single nucleotide variant (nonsense)	CRKX
NIL001286880.1(TTN):C.13535A>C (P.G144512P>R)	TTN	DCM	Uncertain significance	single nucleotide variant (S12 a.a. change : G144512P>R)	CRKX
NIL001286880.1(TTN):C.13711C>G (P.R1404571A>L)	TTN	DCM	Uncertain significance	single nucleotide variant (S127 a.a. change : R1404571A>L)	CRKX
NIL001286880.1(TTN):C.15335G>A (P.G145157(S>P))	TTN	DCM	Uncertain significance	single nucleotide variant (S127 a.a. change : G145157(S>P))	CRKX
NIL001286880.1(TTN):C.16023A (P.I44541E>T)	TTN	DCM	Pathogenic	single nucleotide variant (S4 a.a. change : I44541E>T)	CRKX
NIL001286880.1(TTN):C.16147A (P.L55334(S>N))	TTN	DCM	Uncertain significance	single nucleotide variant (S39 a.a. change : L55334(S>N))	CRKX
NIL001286880.1(TTN):C.16273C>T (P.L55425P>H)	TTN	DCM	Uncertain significance	single nucleotide variant (S423 a.a. change : L55425P>H)	CRKX

Figure 3.4 List for gene variant. Click the gene name on Figure 3.3, get variant list and variant type about the gene.

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
------	-----------------	-------	--------	----------

Variant of *TTN* related to cardiomyopathy

Gene : *TTN*
 Protein : Titin
 Variation : NM_001256850.1(TTN):C.114540DEL(T.P,ASN3818LYSFS)
 Variation Type : deletion frameshift

< Protein Structure >				
Structure	Domain	Function	Post_Translational_Modification	Subcellular_Location
STRAND 4 10	DOUJAN 6 96 Q-like 1	UP_BIND 32164 32182 ATP	CHAIN 1 34350 TTN	
STRAND 15 18	DOUJAN 104 182 Q-like 2	ACT_SITE 32256 32265 Proton acceptor	MOD_RES 263 263 Phosphoserine	
STRAND 23 33	REPEAT 417 462 Z-HOMER 1	BINDING 32207 32207 ATP	MOD_RES 265 265 Phosphoserine	
STRAND 36 45	REPEAT 466 511 Z-HOMER 2		MOD_RES 267 267 Phosphothreonine	
TURIN 48 50	REPEAT 512 554 Z-HOMER 3		MOD_RES 4069 4069 Phosphoserine	
STRAND 55 59	REPEAT 555 600 Z-HOMER 4		MOD_RES 4069 4069 Phosphoserine	
STRAND 62 69	REPEAT 601 648 Z-HOMER 5		MOD_RES 6920 6920 Phosphoserine	
HELIX 72 74	REPEAT 647 691 Z-HOMER 6		MOD_RES 8490 8490 Phosphotyrosine	
STRAND 76 84	REPEAT 692 740 Z-HOMER 7		MOD_RES 9122 9122 Phosphoserine	
STRAND 87 95	DOUJAN 843 1001 Q-like 3		MOD_RES 9203 9203 Phosphoserine	
STRAND 102 105	DOUJAN 1082 1172 Q-like 4		MOD_RES 9207 9207 Phosphothreonine	
STRAND 113 116	DOUJAN 1291 1382 Q-like 5		MOD_RES 11803 11803 Phosphoserine	
STRAND 119 131	DOUJAN 1487 1546 Q-like 6		MOD_RES 12007 12007 Phosphothreonine	
STRAND 134 139	DOUJAN 1556 1646 Q-like 7		MOD_RES 12009 12009 Phosphoserine	
STRAND 142 144	DOUJAN 1703 1793 Q-like 8		MOD_RES 12022 12022 Phosphoserine	
	DOUJAN 32617 32710 Q-like 144			
	DOUJAN 32722 32811 Q-like 145			
	DOUJAN 33301 33391 Q-like 146			
	DOUJAN 33489 33576 Q-like 147			
	DOUJAN 33648 33732 Q-like 148			
	DOUJAN 33779 33867 Q-like 149			
	DOUJAN 33963 34052 Q-like 150			
	DOUJAN 34061 34149 Q-like 151			
	DOUJAN 34256 34344 Q-like 152			
	REGION 253 341 Z-B1			
	REGION 1410 1440 Z-B5			
	COILED 559 561			
	COILED 2058 2062			
	COILED 3462 3467			
	COILED 9534 9577			
	COUPLER 391 436 A-W-HCN			
	COUPLER 453 456 Poly-Thr			
	COUPLER 9500 9503 Poly-Glu			
	COUPLER 9561 9562 Poly-HCN			
	COUPLER 9674 11917 Glu-HCN			
	COUPLER 9674 10059 Glu-HCN			
	COUPLER 10102 10105 Poly-His			
	COUPLER 10211 12032 Poly-HCN			
	COUPLER 33155 33155 Poly-Ser			
	COUPLER 33157 33200 Poly-Ala			
	COUPLER 34102 34244 Ser-HCN			

Copyright 2017 Lab of Computational Biology and Bioinformatics All rights reserved
Seoul National University, Korea +82-2-500-2352

Figure 3.5 Information about the protein structure. It contains information about secondary structure of protein, domain, functional region, PTM, subcellular location.

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
------	-----------------	-------	--------	----------

Search for variant related to cardiomyopathy

- Select the options to investigate variants and click the "Search" button

<Find variants of Dilated cardiomyopathy>

Gene symbol: All Genes Clinical_significance: All significance Variant Type: All types [Search](#)

<Find variants of Hypertrophic cardiomyopathy>

Gene symbol: All Genes Clinical_significance: All significance Variant Type: All types [Search](#)

<Find variants of Restrictive cardiomyopathy>

Gene symbol: All Genes Clinical_significance: All significance Variant Type: All types [Search](#)

<Find variants of Arrhythmogenic right ventricular cardiomyopathy>

Gene symbol: All Genes Clinical_significance: All significance Variant Type: All types [Search](#)

<Find variants of Left ventricular noncompaction cardiomyopathy>

Gene symbol: All Genes Clinical_significance: All significance Variant Type: All types [Search](#)

Copyright 2017 Lab of Computational Biology and Bioinformatics All rights reserved
Seoul National University, Korea *02-2-800-2052

Figure 3.6 Search page. Select condition of options about 'Gene symbol', 'Clinical significance', 'Variant Type'.

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
------	-----------------	-------	--------	----------

Variants and Genes of Dilated cardiomyopathy

Clinical significance : Pathogenic
Total variant : 375
Total genes : 39
Gene set
ABCC9, BAG3, DMD, EYA4, MT-TV, NEXN, PSEN1, MYBPC3, MYH7, SCN5A, TNXB3, TNNT2, TPM1, FKBP, LDB3, ACTN2, RBM20, TTR, CRYAB, DES, DSG2, LAMA4, MYH6, PLN, RAFT1, TNNT1, TCAP, SDHA, DSP, ACTC1, LMNA, TSMF, TNXBK, GATAD1, PRDM16, MYPK, DRAJIS, SGCD, VCL

Variation	Gene	Disease	Clinical_significance	Variation_Type	Protein
ABCC9, 3-BP DEL, 4-BP IN, E138	ABCC9	DCM	Pathogenic	indel	click
BAG3, 1-BP DEL, E52C	BAG3	DCM	Pathogenic	deletion	click
BAG3, E142E	BAG3	DCM	Pathogenic	deletion	click
DMD, 18-BP DEL	DMD	DCM	Pathogenic	deletion	click
DMD, A1471A	DMD	DCM	Pathogenic	insertion	click
DMD, E171E	DMD	DCM	Pathogenic	deletion	click
DMD, I181, G-T, +1	DMD	DCM	Pathogenic	splice site mutation	click
EYA4, 4848-BP DEL	EYA4	DCM	Pathogenic	deletion	click
U58434G	MT-TV	DCM	Pathogenic	mitochondrially encoded variant	click

Figure 3.7 Genetic information that satisfying options of Search page. Show the variations of the condition selected from the figure 3.6. The 'Gene set' could be used for gene panel to genetic screening test.

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
------	-----------------	-------	--------	----------

Download DNA or amino acid sequence
 - Click the gene name to download FASTA file

< Dilated cardiomyopathy >

ABCC9	ACADVL	ACTC1	ACTN2	ANKRD1	BAG3	CALR3	CBRP3	CTNNA3	DEB	DMD	DBC2	DSP	EYA4	FKTN	GATAD1
HCN4	JPH2	JUP	LAMP2	LD33	LMNA	MYBPC3	MYH6	MYH7	MYL2	MYPN	NEBL	NEXN	POLM3	PKP2	PLN
PSEN1	RBM20	RYZR	SCN5A	SCZ	SGCD	TCAP	TMEM43	TMPO	TNNC1	TNNI3	TNNI2	TPM1	TBM	TTN	VCL

< Hypertrophic cardiomyopathy >

ABCC9	ACADVL	ACTC1	ACTN2	ANKRD1	BAG3	CALR3	CBRP3	CTNNA3	DEB	DMD	DBC2	DSP	EYA4	GLA	JPH2
JUP	LAMP2	LD33	LMNA	MYBPC3	MYH6	MYH7	MYL2	MYL3	MYLX2	MYOM	MYOZ2	MYPN	NEBL	NEXN	POLM3
PKP2	PLN	PRKAG2	RBM20	RYZR	SCN5A	TAZ	TCAP	TMEM43	TMPO	TNNC1	TNNI3	TNNI2	TPM1	TTN	VCL

< Restrictive cardiomyopathy >

DSP	MYH7	MYPN	TNNI3	TNNI2
-----	------	------	-------	-------

< Arrhythmogenic right ventricular cardiomyopathy >

ABCC9	DMD	DBC2	DBC2	DSP2	HCN4	JUP	MYOM1	PKP2	RBM20	RYZR	SCN5A	TGFB3	TMEM43	TTN
-------	-----	------	------	------	------	-----	-------	------	-------	------	-------	-------	--------	-----

< Left ventricular noncompaction cardiomyopathy >

ACTC1	ACTN2	CTNNA3	DEB	DMD	DSP	JPH2	JUP	LD33	MYH7	MYPN	NEXN	RYZR	SCN5A	TCAP	TNNI2	TPM1	TTN
-------	-------	--------	-----	-----	-----	------	-----	------	------	------	------	------	-------	------	-------	------	-----

Copyright 2017 Lab of Computational Biology and Bioinformatics All rights reserved
 Seoul National University, Korea +82-2-880-2852

Figure 3.8 Gene list for download about sequence data. All genes that belong to each subtypes of cardiomyopathy.

Cardiomyopathy Genetic Database

Home	Cardiomyopathy?	Genes	Search	Download
------	-----------------	-------	--------	----------

DNA and amino acid sequence data(GenBank) of TTN
 - Click the "Check!!!!" to investigate Definition
 - Click the "FASTA" to download FASTA file of DNA or amino acid

Accession	GI	Keyword	bp.length	Definition	Origin	AAseq
NG_011618	GI:42570063	RefSeq: RefSeqGene.	31688	Check!!!	FASTA	FASTA
NIL001286880	GI:37948564	RefSeq.	106791	Check!!!	FASTA	FASTA
NIL001287880	GI:64294630	RefSeq.	110784	Check!!!	FASTA	FASTA
NIL003379	GI:29154822	RefSeq.	8300	Check!!!	FASTA	FASTA
NIL133378	GI:29154824	RefSeq.	102970	Check!!!	FASTA	FASTA
NIL133379	GI:42580089	RefSeq.	12486	Check!!!	FASTA	FASTA
NIL133432	GI:29154827	RefSeq.	8381	Check!!!	FASTA	FASTA
NIL133437	GI:64294632	RefSeq.	83788	Check!!!	FASTA	FASTA

Copyright 2017 Lab of Computational Biology and Bioinformatics All rights reserved
 Seoul National University, Korea +82-2-880-2852

Figure 3.9 Sequence data about selected gene. Shows the sequence information of the gene selected in the figure 3.8. Click the 'FASTA' button to download fasta format file.

Table 3.1 Gene list through the gene searching system

Subtypes	Genes
DCM	ABCC9, ACTC1, ACTN2, BAG3, CRYAB, DES, DMD, DNAJB6, DSG2, DSP, EYA4, FKTN, GATAD1, LAMA4, LDB3, LMNA, MYBPC3, MYH6, MYH7, MYPN, NEXN, PLN, PRDM16, PSEN1, RAF1, RBM20, SCN5A, SCO2, SDHA, SGCD, TCAP, TNNC1, TNNI3, TNNT2, TPM1, TSFM, TTN, VCL
HCM	ACTC1, ACTN2, CAV3, CSRP3, DES, FHL1, JPH2, LAMP2, MYBPC3, MYH6, MYH7, MYL2, MYLK2, MYO6, MYOZ2, NEXN, PLN, PMPCA, PRKAG2, RAF1, TNNC1, TNNI3, TNNT2, TPM1, TTN, VCL
RCM	MYH7, MYPN, TNNI3, TNNT2
ARVC	DSC2, DSG2, DSP, JUP, PKP2, RYR2, TGFB3, TMEM43
LVNC	MYH7, RYR2, SCN5A, TTN

Genes with mutations whose clinical significance is 'pathogenic' or 'likely pathogenic'. The genes could be used for gene panel to genetic screening test.

3.2 마이크로어레이 데이터를 이용한 차별발현 유전자 분석

본 연구에서는 다양한 원인과 종류에 의한 심근증 환자들 중 심부전의 예후를 보이는 사람들의 심장에서 발현되는 유전자들의 정보를 이용하여 차별발현 유전자 분석을 수행 하였다. 그 후 각 comparison에서 차별발현 되는 유전자들에 모두 속하는 유전자들을 선별하여 확장성 심근증, 허혈성 심근증, 비대성 심근증 환자들에서 공통으로 발현이 조절되는 유전자들의 목록을 만들었다. 본 연구에서 이용한 마이크로어레이 데이터들 중 확장성 심근증, 허혈성 심근증, 비대성 심근증 그룹의 공통점은 세 그룹 다 심부전이란 예후가 나타났다는 것이다. 즉, 최종적으로 얻은 목록의 유전자들은 심근증 환자들에게서 심부전의 예후가 나타나게 하는 경로의 유전자들이라 생각 할 수 있다.

<Table 3.2>에는 정상인의 데이터와 각각의 다른 원인에 따른 심근증 환자들의 데이터를 비교하여 차별발현 되는 유전자 및 각 실험군에서 공통으로 조절되는 유전자들의 수를 나타내었다. 실험군인 idiopathic dilated cardiomyopathy 환자들과 대조군인 정상인들을 비교한 comparison1에서는 발현량이 증가하는 유전자가 310개, 감소하는 유전자가 314개, 총 624개의 DEGs가 관찰되었다. 실험군이 ischemic cardiomyopathy 환자들인 comparison2 에서는 발현량이 증가하는 유전자가 403개, 감소하는 유전자가 161개, 총 564개의 DEGs가 관찰되었다. 실험군이 hypertrophic cardiomyopathy 환자들인 comparison3 에서는 발현량이 증가하는 유전자가 240개 감소하는 유전자가 123개, 총 363개의 DEGs가 관찰되었다.

이렇게 얻어진 DEGs중 확장성 심근증, 허혈성 심근증, 비대성 심근증 환자들에서 공통으로 발현이 조절되는 유전자들은 총 73개

가 관찰 되었으며, 발현량이 증가하는 유전자는 55개, 발현량이 감소하는 유전자는 18개가 관찰되었다. 위의 공통적으로 조절되는 유전자의 목록은 <Table 3.3>에 나타내었다.

심근증으로 인한 심부전에서 발현량이 조절되는 유전자들인 comparison4의 유전자들의 기능을 알아보기 위해 DAVID (the Database for Annotation, Visualization and Integrated Discovery)를 이용해 functional annotation을 실시한다. DAVID는 웹 기반의 생물학적 해석 시스템으로, 대규모의 유전자 및 단백질 목록에서 생물학적 의미를 체계적으로 추출하기 위한 도구이다(Huang et al., 2009). 이를 이용하여 마이크로어레이 분석을 통해 얻은 결과들을 분석 할 수 있는데, “Functional Annotation Chart”기능을 통해 Gene Ontology, General Annotation등의 여러 데이터베이스에 수록된 정보들을 얻을 수 있으며 이에 대한 p-value와 adjusted p-value를 확인 할 수 있다. David를 통해 전체 Co-EGs와 up-regulated 및 down-regulated되는 유전자들의 Functional annotation을 수행하면, 전체 Co-EGs에 대한 주요한 내용은 ‘UP_KEYWORDS’ 카테고리의 Secreted, Signal, Collagen, ‘GOTERM_CC_DIRECT’ 카테고리의 proteinaceous extra -cellular matrix, extracellular region, extracellular space, collagen trimer, extracellular matrix 등이 있다. Up-regulated되는 유전자들의 주요한 내용은 ‘UP_KEYWORDS’카테고리의 Secreted, Extracellular matrix, ‘GOTERM_CC_DIRECT’카테고리의 extracellular region, extracellular space, ‘GOTERM_BP_DIRECT’카테고리의 extra-cellular matrix organization, axonogenesis등이 있다. Down-regulated되는 유전자들의 주요한 내용은 ‘UP_KEYWORDS’카테고리의 Actin-binding, Myosin, ‘GOTERM_MF_DIRECT’카테고리의 actin filament binding, ‘GOTERM_CC_DIRECT’카테고리의 sarcomere, myosin complex등이 있다 <Table 3.4>.

Table 3.2 Quantity of DEGs and Co-EGs.

Dataset	Toatal	The number of DEGs	
		Up - regulated	Down – regulated
comparison 1	624	310	314
comparison 2	564	403	161
comparison 3	363	240	123

Dataset	Toatal	The number of Co-EGs	
		Up - regulated	Down – regulated
comparison 4	73	55	18

DEGs : differentially expressed genes

Co-EGs : Co-expressed genes

Number of genes selected by differentially expressed gene analysis and Co-expressed gene analysis.

Table 3.3 Co-regulated genes of comparison4

Regulated	Co-regulated genes
Up (55)	ARRDC3, ASPN, C1QTNF2, C1QTNF7, CCNG2, CD47, CEP126, CNTN4, COL14A1, COLEC12, CRISPLD1, CTDSPL, CYS1, ECM2, FAM188A, FBXL20, FNDC1, FRZB, GLT8D2, HAPLN1, HNMT, IFI44L, IGF1, ITGBL1, ITIH5, KIAA1211, LRRC17, LSP1, LUM, MATN2, MBP, MEOX2, MNS1, N6AMT1, NDC80, NRK, OGN, OIP5-AS1, OMD, PCSK5, PENK, POSTN, PTN, RAPGEF6, RERGL, SCG5, SCN2B, SFRP4, SMOC2, SPATA18, SSPN, SULF1, TACC1, ZBTB20, ZNF704
Down (18)	AGO2, ARMC8, CCL2, EHD4, FCN3, FLNC, GTPBP1, IL1RL1, METTL7B, MYH6, MYO16, NAMPT, NME7, PYROXD1, SERPINA3, SLC4A2, SLCO4A1, SYNE1

The above genes show the same amount of change in the three experimental groups. This makes it possible to see that these genes play important roles in the development of heart failure in cardiomyopathy.

Table 3.4 Major components of Functional annotation

Category	Term	%	p-value	adjusted p-value
< All Co-EGs >				
UP_KEYWORDS	Secreted	41	2.70E-12	4.27E-10
GOTERM_CC_DIRECT	proteinaceous extracellular matrix	15	5.49E-08	6.32E-06
UP_SEQ_FEATURE	signal peptide	44	8.21E-08	3.23E-05
UP_KEYWORDS	Extracellular matrix	12	2.94E-06	2.33E-04
UP_KEYWORDS	Signal	44	6.28E-06	3.30E-04
GOTERM_CC_DIRECT	extracellular region	26	1.71E-05	9.81E-04
GOTERM_CC_DIRECT	extracellular space	23	2.70E-05	0.00103
GOTERM_CC_DIRECT	collagen trimer	7	3.87E-04	0.01106
UP_KEYWORDS	Collagen	7	3.68E-04	0.01442
GOTERM_CC_DIRECT	extracellular matrix	10	8.21E-04	0.01871
UP_KEYWORDS	Leucine-rich repeat	10	7.13E-04	0.02227
UP_KEYWORDS	Disulfide bond	33	8.60E-04	0.02240
<Up-regulated Co-EGs >				
UP_KEYWORDS	Secreted	45	1.44E-11	1.73E-09
GOTERM_CC_DIRECT	proteinaceous extracellular matrix	18	5.96E-08	5.12E-06
UP_SEQ_FEATURE	signal peptide	49	4.98E-08	1.24E-05
UP_KEYWORDS	Extracellular matrix	16	2.98E-07	1.79E-05
UP_KEYWORDS	Signal	49	2.49E-06	9.94E-05
GOTERM_CC_DIRECT	extracellular region	29	2.82E-05	0.00121
GOTERM_CC_DIRECT	extracellular space	25	7.73E-05	0.00221
GOTERM_CC_DIRECT	extracellular matrix	13	1.90E-04	0.00406
UP_KEYWORDS	Leucine-rich repeat	13	1.43E-04	0.00429
GOTERM_BP_DIRECT	extracellular matrix organization	13	2.16E-05	0.00795

Table 3.4 Major components of Functional annotation (cont.)

Category	Term	%	p-value	adjusted p-value
<Down-regulated Co-EGs >				
UP_KEYWORDS	Actin-binding	22	0.00138	-
GOTERM_MF_DIRECT	actin filament binding	17	0.00764	-
GOTERM_BP_DIRECT	positive regulation of nitric-oxide synthase biosynthetic process	11	0.01232	-
UP_SEQ_FEATURE	domain:Actin-binding	11	0.01681	-
INTERPRO	Actinin-type, actin-binding, conserved site	11	0.02086	-
GOTERM_CC_DIRECT	sarcomere	11	0.03163	-
INTERPRO	Myosin head, motor domain	11	0.03603	-
GOTERM_CC_DIRECT	myosin complex	11	0.04038	-
UP_KEYWORDS	Myosin	11	0.04131	-
SMART	SM00242:MYSc	11	0.04291	-

The genes whose expression levels increase are mainly related to Extracellular matrix, and the genes whose expression levels decrease are mainly related to contraction of the heart muscle. Since the number of genes to be down-regulated in Co-EGs is not large, the significance level is indicated by p-value.

3.3 유전자 네트워크 분석

차별발현 유전자 분석으로 얻은 comparison4의 유전자 목록으로 네트워크를 구성하여 네트워크의 밀도 중심성에 대한 분석을 실시하였다. 먼저 네트워크의 전체적인 형태를 살펴보기 위해 피어슨 상관계수의 절대값이 0.5, 0.7이상인 기준이 되는 가중치가 없는 무방향 네트워크를 각각 구성하였으며 이는 <Figure 3.10>을 통해 확인할 수 있다. 피어슨 상관계수의 절대값이 0.7이상인 interaction으로 구성된 네트워크의 전체적인 형상을 보면 크게 두 개의 서브네트워크가 관찰되는데 하나의 네트워크는 up-regulated되는 유전자들로 구성되어있고 다른 하나는 up-regulated되는 유전자들과 down-regulated되는 유전자들로 함께 구성되어있는데, 이는 NME7과 OIP5-AS1 유전자들의 연결선으로 인해 up-regulated되는 부분과 down-regulated되는 부분으로 나뉜다. 피어슨 상관계수의 절대값이 0.5이상인 interaction으로 네트워크를 구성하면 두 서브네트워크 사이에 연결선이 형성된다. 위의 결과를 보면 comparison4의 유전자들은 up-regulated및 down-regulated되는 유전자들끼리 서브네트워크를 형성하여 그 안에서 서로 양의 상관관계로 발현이 조절되며 각각의 서브네트워크 사이에는 음의 상관관계로 상대 네트워크의 발현을 조절하는 몇몇의 허브 유전자들이 있을 것이라 생각된다.

네트워크의 밀도는 '가능한 연결선의 수' 대비 '총 연결선의 수'로 정의되는데, comparison4의 유전자를 노드로 피어슨 상관계수의 절대값이 0.5이상인 interaction을 연결선으로 구성한 가중치가 없는 무방향 네트워크의 밀도는 0.232로 나타난다.

다음으로 피어슨 상관계수의 절대값이 0.5 이상인 interaction을 연결선으로 구성한 무방향 네트워크의 소셜네트워크 중심성 분석을 실시하였다. 각 노드에 연결된 연결선의 수를 나타내는 연결정도 중

심성은 ECM2가 제일 높게 나오고 SMOC2, OGN, SULF1, FRZB, LUM의 순서로 높게 나온다. 하나의 노드에서 각각의 노드 사이의 평균 최단거리의 역수로 나타내어 지는 근접중심성은 SMOC2가 제일 높게 나오고 ECM2, SULF1, OGN, NRK, SPATA18의 순서로 높게 나온다. 노드와 노드를 연결해 주는 정도를 나타내는 중개중심성은 HNMT가 제일 높게 나오고 PTN, NAMPT, FAM188A, ITIH5, MYH6의 순서로 높게 나온다. Network2의 중심성 분석 결과는 <Table 3.5>에서 확인 할 수 있다. 위의 중심성 분석에서 연결정도 중심성과 근접중심성은 노드의 수가 많은 up-regulated되는 유전자들로 구성된 서브네트워크에 속해있는 유전자들이 높게 나왔지만 중개중심성은 서브네트워크가 인접한 부분에 위치하여 서브네트워크 간에 연결선을 형성하는데 중요한 역할을 하는 유전자들이 높게 나왔다. 이렇게 얻은 중심성의 결과들을 하나로 지표로 통합하기 위해 하나의 항목이 각각의 지표에서 차지하는 비율을 구하여 세가지 지표를 표준화 하고, 각 유전자가 가지는 세 지표의 평균 값을 구한다. Network2의 통합적인 중심성 정도는 HNMT, PTN, ITIH5, SMOC2, SULF1, NAMPT, FAM188A, ECM2, OIP5-AS1, C1QTNF7가 높게 나왔고, <Table 3.6>을 통해 그 기능을 알 수 있다.

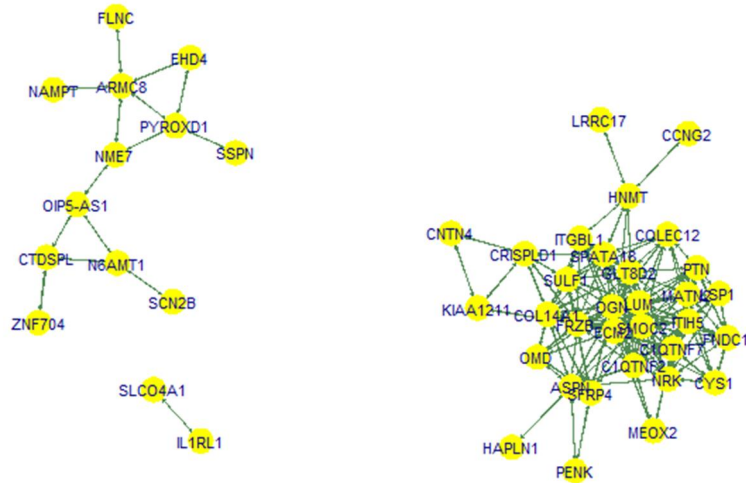
다음으로는 2부 네트워크를 구성하여 comparison4로 구성된 네트워크와 확장성 및 비대성 심근증 환자에서 선천적인 변이를 보이는 유전자들과의 관계를 알아보았고 이를 통해 심근증의 예후 중심부전과 관련이 있는 유전자들을 유추 하였다. 먼저 확장성 심근증과 관련이 있는 유전자들과 comparison4간에 피어슨 상관계수의 절대값이 0.6이상인 interaction을 연결선으로 구성한 가중치가 없는 무방향 네트워크를 구성하고 중심성 분석을 수행하였는데 구성된 네트워크는 <Figure 3.11>을 통해 확인 할 수 있다. 선천적으로 확장성 심근증과 관계가 있는 유전자들을 기준으로 한 2부네트워크의 중심

성 분석에서 연결정도 중심성은 VCL이 제일 높게 나오고 DSG2, DSP, ILK, MYPN, NEXN, PKP2의 순서로 높게 나온다. 근접중심성은 VCL이 제일 높게 나오고 DSG2, DSP, PKP2, MYPN, NEXN, ILK의 순서로 높게 나온다. 중개중심성은 PSEN1이 제일 높게 나오고 ILK, RBM20, VCL, LAMP2, MYPN, FKTN의 순서로 높게 나온다. 같은 2부 네트워크에서 comparison4를 중심으로 중심성 분석을 하면 연결정도 중심성은 OIP5-AS1이 제일 높게 나오고 FLNC, EHD4, MYH6, SCN2B, SSPN의 순서로 높게 나온다. 근접중심성은 OIP5-AS1이 제일 높게 나오고 FLNC, EHD4, MYH6, SCN2B, SSPN의 순서로 높게 나온다. 중개 중심성은 OIP5-AS1이 제일 높게 나오고 FLNC, MYH6, SCN2B, SLC4A2, EHD4의 순서로 높게 나온다. Network3의 중심성 분석 결과는 <Table 3.7>과 <Table 3.8>을 통해 확인 할 수 있다. Network3의 통합적인 중심성 정도는 확장성 심근증과 관계가 있는 유전자들 기준으로는 PSEN1, ILK, VCL, RBM20, MYPN, LAMP2, PKD1, FKTN, DSG2, DSP의 순으로 높게 나오고, comparison4 기준으로는 OIP5-AS1, FLNC, MYH6, SCN2B, EHD4, ARMC8, SSPN, NME7, PYROXD1, SLC4A2의 순으로 높게 나온다. 위 유전자들의 기능은 <Table 3.9>를 통해 확인 할 수 있다.

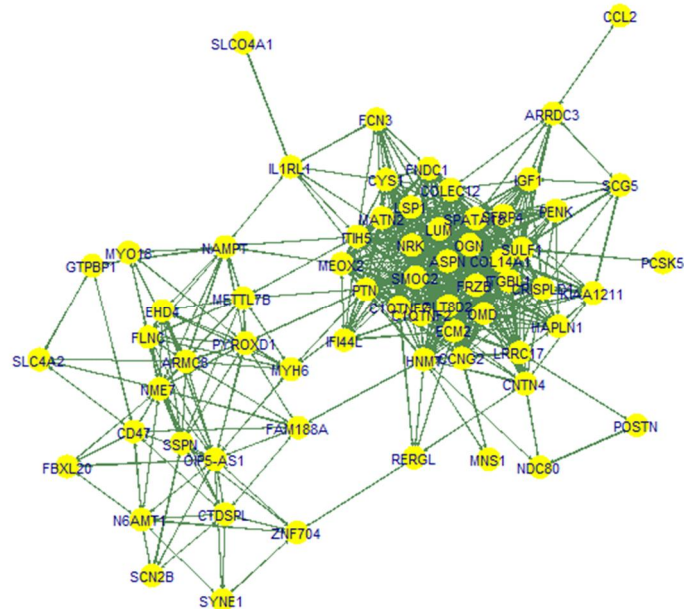
두 번째로 비대성 심근증과 관련이 있는 유전자들과 comparison4 간에 피어슨 상관계수의 절대값이 0.6이상인 interaction을 연결선으로 구성한 가중치가 없는 무방향 네트워크를 구성하고 중심성 분석을 수행하였는데 구성된 네트워크는 <Figure 3.12>를 통해 확인 할 수 있다. 비대성 심근증의 원인이 되는 유전자를 중심으로 중심성 분석을 하면 연결정도 중심성은 VCL이 제일 높게 나오고 MYPN, NEXN, PRKAG2의 순서로 높게 나온다. 근접중심성은 VCL이 제일 높게 나오고 MYPN, PLN, NEXN, PRKAG2, MYOZ2의 순서로 높게 나온다. 중개중심성은 PLN이 제일 높게 나오고 MYOZ2,

LAMP2, VCL, MYPN, PRKAG2, TTN의 순서로 높게 나온다. 같은 2부 네트워크에서 comparison4를 중심으로 중심성 분석을 하면 연결정도 중심성은 PYROXD1이 제일 높게 나오고 FLNC, SSPN, ARMC8, MYO16, SCN2B의 순서로 높게 나온다. 근접중심성은 PYROXD1이 제일 높게 나오고 SSPN, FLNC, ARMC8, MYO16, SCN2B의 순서로 높게 나온다. 중개중심성은 PYROXD1이 제일 높게 나오고 SCN2B, SSPN, FLNC, N6AMT1, ARMC8의 순서로 높게 나온다. Network4의 중심성분석 결과는 <Table 3.10>과 <Table 3.11>을 통해 확인 할 수 있다. Network4의 통합적인 중심성 정도는 확장성 심근증과 관계가 있는 유전자들 기준으로는 PLN, VCL, MYOZ2, LAMP2, MYPN, PRKAG2, NEXN, TTN의 순으로 높게 나오고, comparison4 기준으로는 PYROXD1, SSPN, SCN2B, FLNC, ARMC8, MYO16의 순서로 높게 나온다. 위 유전자들의 기능은 <Table 3.12>를 통해 확인 할 수 있다.

a)



b)



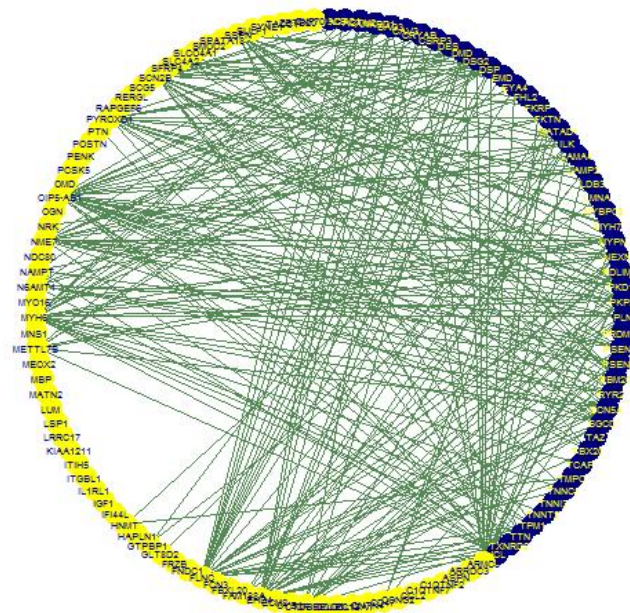


Figure 3.11 Two-mode network between Comparison4 genes and germline variant of dilated cardiomyopathy. Yellow nodes are Co-EGs and blue nodes are DCM genes.

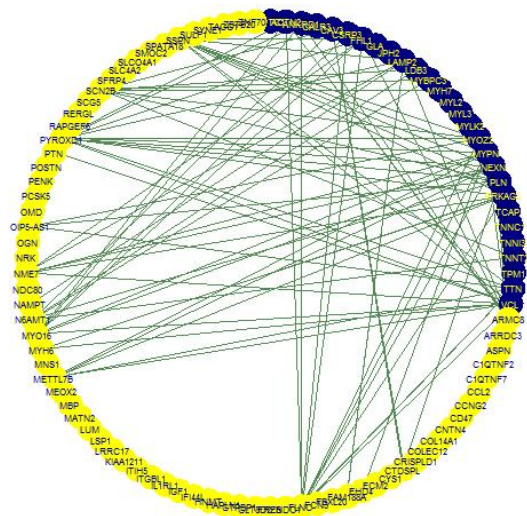


Figure 3.12 Two-mode network between Comparison4 genes and germline variant of hypertrophic cardiomyopathy. Yellow nodes are Co-EGs and blue nodes are HCM genes.

Table 3.5 Centrality analysis of network2

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Centrality	Gene	Centrality
ECM2	0.986	SMOC2	0.684	HNMT	0.232
SMOC2	0.986	ECM2	0.673	PTN	0.187
OGN	0.957	SULF1	0.663	NAMPT	0.151
SULF1	0.957	OGN	0.661	FAM188A	0.145
FRZB	0.928	NRK	0.656	ITIH5	0.143
LUM	0.928	SPATA18	0.656	MYH6	0.108
NRK	0.928	FRZB	0.653	SMOC2	0.107
SPATA18	0.928	LUM	0.653	OIP5-AS1	0.101
COL14A1	0.899	HNMT	0.652	SULF1	0.098
ITGBL1	0.899	PTN	0.652	NME7	0.082
ASPN	0.870	ITGBL1	0.649	METTL7B	0.082
C1QTNF7	0.870	ITIH5	0.649	PYROXD1	0.076
GLT8D2	0.870	C1QTNF7	0.646	RERGL	0.068
OMD	0.870	COL14A1	0.646	LRRC17	0.064
C1QTNF2	0.841	ASPN	0.639	ECM2	0.063
COLEC12	0.841	GLT8D2	0.639	IL1RL1	0.063
HNMT	0.841	OMD	0.639	ARRDC3	0.055
ITIH5	0.841	C1QTNF2	0.632	MEOX2	0.052
PTN	0.841	COLEC12	0.632	C1QTNF7	0.050
SFRP4	0.841	SFRP4	0.627	ZNF704	0.043
CRISPLD1	0.783	LSP1	0.610	SPATA18	0.035
LSP1	0.754	MATN2	0.607	NRK	0.034
MATN2	0.754	CRISPLD1	0.606	ARMC8	0.033
FNDC1	0.696	MEOX2	0.599	ITGBL1	0.032
CCNG2	0.667	CCNG2	0.588	COLEC12	0.028
CYS1	0.667	FNDC1	0.586	SSPN	0.025
LRRC17	0.667	CYS1	0.579	COL14A1	0.023
MEOX2	0.667	LRRC17	0.577	CD47	0.022
HAPLN1	0.522	HAPLN1	0.535	CTDSPL	0.019
IGF1	0.464	IGF1	0.523	GLT8D2	0.019

Table 3.5 Centrality analysis of network2 (cont.)

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Centrality	Gene	Centrality
PENK	0.271	PENK	0.516	OGN	0.018
NME7	0.271	PYROXD1	0.514	SFRP4	0.017
OIP5-AS1	0.237	NAMPT	0.505	LUM	0.015
KIAA1211	0.237	CNTN4	0.504	CRISPLD1	0.013
CNTN4	0.220	METTL7B	0.500	FRZB	0.013
ARMC8	0.186	KIAA1211	0.498	CYS1	0.011
PYROXD1	0.169	MYH6	0.498	OMD	0.010
FCN3	0.153	FAM188A	0.481	IGF1	0.010
NAMPT	0.153	FCN3	0.481	CNTN4	0.009
FLNC	0.153	IFI44L	0.476	LSP1	0.008
IFI44L	0.136	NME7	0.473	N6AMT1	0.008
METTL7B	0.136	OIP5-AS1	0.473	ASPN	0.008
SSPN	0.136	IL1RL1	0.467	C1QTNF2	0.007
CTDSPL	0.119	ARMC8	0.457	GTPBP1	0.006
EHD4	0.119	RERGL	0.452	NDC80	0.006
CD47	0.085	FLNC	0.436	FLNC	0.005
MYH6	0.085	ARRDC3	0.430	MATN2	0.005
N6AMT1	0.068	EHD4	0.429	EHD4	0.004
ARRDC3	0.068	SCG5	0.425	FCN3	0.003
FAM188A	0.068	SSPN	0.421	KIAA1211	0.003
IL1RL1	0.068	CTDSPL	0.415	FNDC1	0.002
SCG5	0.051	MNS1	0.411	CCNG2	0.002
ZNF704	0.051	CD47	0.405	SCG5	0.001
MYO16	0.051	ZNF704	0.399	SLC4A2	0.001
RERGL	0.034	NDC80	0.399	HAPLN1	0.001
FBXL20	0.034	MYO16	0.381		
SCN2B	0.017	PCSK5	0.377		
SLC4A2	0.017	N6AMT1	0.376		
SYNE1	0.017	GTPBP1	0.360		
GTPBP1	0.017	POSTN	0.353		

The centrality values are the normalized values and the genes are listed in descending order of centrality.

Table 3.6 Important genes of network2 (high ranked 10 genes)

Gene	Protein	Function
HNMT	Histamine N-methyltransferase	Plays an important role in degrading histamine and in regulating the airway response to histamine
PTN	Pleiotrophin	Secreted growth factor that induces neurite outgrowth and which is mitogenic cells
ITIH5	Inter-alpha-trypsin inhibitor heavy chain H5	May act as a tumor suppressor
SMOC2	SPARC-related modular calcium-binding protein 2	Promotes matrix assembly and cell adhesiveness
SULF1	Extracellular sulfatase Sulf-1	Exhibits arylsulfatase activity and highly specific endoglucosamine-6-sulfatase activity
NAMPT	Nicotinamide Phosphoribosyltransferase	Catalyzes the condensation of nicotinamide with 5-phosphoribosyl-1-pyrophosphate to yield nicotinamide mononucleotide
FAM188A	Family With Sequence Similarity 188 Member A	contains a caspase-associated recruitment domain and may function in apoptosis
ECM2	Extracellular matrix protein 2	Promotes matrix assembly and cell adhesiveness
OIP5-AS1	OIP5 Antisense RNA 1 (RNA Gene)	non-coding RNA
C1QTNF7	Complement C1q tumor necrosis factor-related protein 7	An important paralog of this gene is C1QTNF2

Highly centrally located genes are in a position to connect subnetworks.

Table 3.7 Two-mode network centrality analysis of network3 about germline variant of dilated cardiomyopathy (high ranked 30 genes)

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Centrality	Gene	Centrality
VCL	0.324	VCL	0.292	PSEN1	0.143
DSG2	0.294	DSG2	0.290	ILK	0.124
DSP	0.294	DSP	0.290	RBM20	0.120
ILK	0.294	PKP2	0.288	VCL	0.108
MYPN	0.294	MYPN	0.286	LAMP2	0.094
NEXN	0.294	NEXN	0.286	MYPN	0.087
PKP2	0.294	ILK	0.285	FKTN	0.079
DMD	0.265	PLN	0.282	PKD1	0.065
PLN	0.265	SGCD	0.282	DSG2	0.049
SGCD	0.265	DMD	0.280	DSP	0.049
LAMP2	0.235	RYR2	0.271	PKP2	0.047
PKD1	0.235	SCN5A	0.271	NEXN	0.039
PSEN1	0.235	PSEN1	0.266	PLN	0.038
RBM20	0.206	TTN	0.265	SGCD	0.036
RYR2	0.206	CRYAB	0.263	GATAD1	0.035
SCN5A	0.206	LAMP2	0.263	DMD	0.034
BAG3	0.176	PKD1	0.263	EMD	0.028
CRYAB	0.176	TXNRD2	0.261	LMNA	0.028
FKTN	0.176	ACTN2	0.260	RYR2	0.024
TTN	0.176	TPM1	0.260	SCN5A	0.022
TXNRD2	0.176	RBM20	0.254	TMPO	0.021
ACTN2	0.147	FKTN	0.250	CRYAB	0.017
PSEN2	0.147	BAG3	0.248	TTN	0.016
TPM1	0.147	CAV3	0.237	TXNRD2	0.016
CAV3	0.118	PSEN2	0.236	BAG3	0.015
TBX20	0.118	EYA4	0.233	ACTN2	0.013
ACTC1	0.088	ACTC1	0.232	TPM1	0.013
ANKRD1	0.088	TNNC1	0.232	PSEN2	0.006
EYA4	0.088	FHL2	0.231	CAV3	0.006
FHL2	0.088	TBX20	0.231	EYA4	0.005

The centrality values are the normalized values and the genes are listed in descending order of centrality.

Table 3.8 Two-mode network centrality analysis of network3 about comparison4 genes

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Centrality	Gene	Centrality
OIP5-AS1	0.765	OIP5-AS1	0.356	OIP5-AS1	0.335
FLNC	0.735	FLNC	0.339	FLNC	0.277
EHD4	0.676	EHD4	0.333	MYH6	0.235
MYH6	0.647	MYH6	0.317	SCN2B	0.216
SCN2B	0.588	SCN2B	0.303	SLC4A2	0.173
SSPN	0.500	SSPN	0.303	EHD4	0.150
NME7	0.471	NME7	0.295	FAM188A	0.133
ARMC8	0.441	ARMC8	0.290	ARMC8	0.132
PYROXD1	0.441	PYROXD1	0.290	SSPN	0.096
CTDSPL	0.324	CTDSPL	0.258	NME7	0.094
N6AMT1	0.265	N6AMT1	0.244	PYROXD1	0.081
METTL7B	0.206	METTL7B	0.230	CTDSPL	0.037
SLC4A2	0.176	SLC4A2	0.225	CD47	0.030
FAM188A	0.147	NAMPT	0.217	N6AMT1	0.025
NAMPT	0.118	FAM188A	0.207	NAMPT	0.006
CD47	0.088	CD47	0.190	METTL7B	0.006
GTPBP1	0.059	PTN	0.187	HNMT	0.001
HNMT	0.059	MYO16	0.185		
GLT8D2	0.029	SYNE1	0.171		
MYO16	0.029	ZNF704	0.171		
NRK	0.029	GTPBP1	0.131		
PTN	0.029	HNMT	0.128		
SYNE1	0.029	GLT8D2	0.013		
ZNF704	0.029	NRK	0.013		

The centrality values are the normalized values and the genes are listed in descending order of centrality.

Table 3.9 Important genes of network3

Gene	Protein	Function
< Genes associated with dilated cardiomyopathy >		
PSEN1	Presenilin-1	Catalytic subunit of the gamma-secretase complex
ILK	Integrin-linked protein kinase	Receptor-proximal protein kinase regulating integrin-mediated signal transduction
VCL	Vinculin	Actin filament binding protein involved in cell-matrix adhesion and cell-cell adhesion
RBM20	RNA-binding protein 20	Regulates splicing of TTN
MYPN	Myopalladin	Component of the sarcomere that tethers together nebulin and nebulin to alpha-actinin
LAMP2	Lysosome-associated membrane glycoprotein 2	Plays an important role in chaperone-mediated autophagy
PKD1	Polycystin-1	Involved in renal tubulogenesis
< Genes of comparison4 >		
OIP5-AS1	OIP5 Antisense RNA 1 (RNA Gene)	non-coding RNA
FLNC	Filamin-C	Muscle-specific filamin, which plays a central role in muscle cells
MYH6	Myosin-6	Muscle contraction
SCN2B	Sodium channel protein type 1 subunit alpha	Mediates the voltage-dependent sodium ion permeability of excitable membranes
EHD4	EH domain-containing protein 4	ATP- and membrane-binding protein
ARMC8	Armadillo repeat-containing protein 8	Among its related pathways are Immune System
SSPN	Sarcospan	Component of the dystrophin-glycoprotein complex
NME7	Nucleoside diphosphate kinase 7	Major role in the synthesis of nucleoside triphosphates other than ATP
PYROXD1	Pyridine nucleotide-disulfide oxidoreductase domain-containing protein 1	Involved in cellular response to oxidative stress

Among the comparison4, there are concentrated lines in the above genes.

Table 3.10 Two-mode network centrality analysis of network4 about germline variant of hypertrophic cardiomyopathy

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Centrality	Centrality	Gene
VCL	0.563	VCL	0.186	PLN	0.502
MYPN	0.500	MYPN	0.180	MYOZ2	0.308
NEXN	0.438	PLN	0.176	LAMP2	0.261
PRKAG2	0.438	NEXN	0.173	VCL	0.251
FHL1	0.375	PRKAG2	0.173	MYPN	0.189
LAMP2	0.375	MYOZ2	0.172	PRKAG2	0.097
PLN	0.375	FHL1	0.166	TTN	0.077
TTN	0.375	TTN	0.166	NEXN	0.064
ACTN2	0.313	LAMP2	0.163	FHL1	0.044
MYOZ2	0.313	ACTN2	0.159	ACTN2	0.028
ANKRD1	0.250	ANKRD1	0.152	ANKRD1	0.016
CAV3	0.188	CSRP3	0.128	CAV3	0.016
CSRP3	0.125	CAV3	0.120	LDB3	0.002
LDB3	0.125	LDB3	0.113	MYBPC3	0.002
MYBPC3	0.125	MYBPC3	0.113	TNNT2	0.002
TNNT2	0.125	TNNT2	0.113	CSRP3	0.002
MYH7	0.063	MYH7	0.106		

The centrality values are the normalized values and the genes are listed in descending order of centrality.

Table 3.11 Two-mode network centrality analysis of network4 about comparison4 genes

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Centrality	Centrality	Gene
PYROXD1	0.688	PYROXD1	0.211	PYROXD1	0.468
FLNC	0.625	SSPN	0.204	SCN2B	0.430
SSPN	0.625	FLNC	0.194	SSPN	0.405
ARMC8	0.563	ARMC8	0.183	FLNC	0.223
MYO16	0.438	MYO16	0.162	N6AMT1	0.121
SCN2B	0.438	SCN2B	0.161	ARMC8	0.107
NME7	0.375	NME7	0.160	MYO16	0.061
METTL7B	0.313	METTL7B	0.150	NME7	0.049
N6AMT1	0.313	N6AMT1	0.143	OIP5-AS1	0.022
OIP5-AS1	0.188	OIP5-AS1	0.138	CTDSPL	0.021
CTDSPL	0.125	MYH6	0.125	METTL7B	0.020
MYH6	0.125	CTDSPL	0.122	MYH6	0.001
EHD4	0.063	PTN	0.118		
FAM188A	0.063	EHD4	0.115		
NAMPT	0.063	FAM188A	0.109		
PTN	0.063	NAMPT	0.109		

The centrality values are the normalized values and the genes are listed in descending order of centrality.

Table 3.12 Important genes of network4

Gene	Protein	Function
< Genes associated with hypertrophic cardiomyopathy >		
PLN	Cardiac phospholamban	Reversibly inhibits the activity of ATP2A2 in cardiac sarcoplasmic reticulum
VCL	Vinculin	Actin filament binding protein involved in cell-matrix adhesion and cell-cell adhesion
MYOZ2	Myozenin-2	Modulation of calcineurin signaling
LAMP2	Lysosome-associated membrane glycoprotein 2	Plays an important role in chaperone-mediated autophagy
MYPN	Myopalladin	Component of the sarcomere that tethers together nebulin and nebulin to alpha-actinin
PRKAG2	5'-AMP-activated protein kinase subunit gamma-2	Regulating cellular energy metabolism
NEXN	Nexilin	Involved in regulating cell migration through association with the actin cytoskeleton
TTN	Titin	Key component in the assembly and functioning of vertebrate striated muscles
< Genes of comparison4 >		
PYROXD1	Pyridine nucleotide-disulfide oxidoreductase domain-containing protein 1	Involved in cellular response to oxidative stress
SSPN	Sarcospan	Component of the dystrophin-glycoprotein complex
SCN2B	Sodium channel protein type 1 subunit alpha	Mediates the voltage-dependent sodium ion permeability of excitable membranes
FLNC	Filamin-C	Muscle-specific filamin, which plays a central role in muscle cells
ARMC8	Armadillo repeat-containing protein 8	Among its related pathways are Immune System
MYO16	Unconventional myosin-XVI	highly divergent tails are presumed to bind to membranous compartments, which would be moved relative to actin filaments

Among the comparison4, there are concentrated lines in the above genes.

3.4 단백질 네트워크 분석

Comparison4와 확장성 심근증의 원인이 되는 유전자간의 2부 단백질 네트워크에서 확장성 심근증의 원인이 되는 유전자를 중심으로 중심성 분석을 하면 연결정도 중심성은 ACTC1이 제일 높게 나오고 DMD, TCAP의 순서로 높게 나온다. 근접중심성은 ACTC1이 제일 높게 나오고 TCAP, DES, LDB3, VCL의 순서로 높게 나온다. 중개중심성은 ACTC1이 제일 높게 나오고 DMD, TCAP, BAG3, SCN5A의 순서로 높게 나온다. Network5의 중심성 분석 결과는 <Table 3.13>을 통해 확인 할 수 있다. 이를 종합해 보면 ACTC1, DMD, TCAP, SCN5A, BAG3의 순서로 중심도가 높게 나온다. Comparison4와 비대성 심근증의 원인이 되는 유전자간의 2부 단백질 네트워크에서 비대성 심근증의 원인이 되는 유전자를 중심으로 중심성 분석을 하면 연결정도 중심성은 ACTC1이 제일 높게 나오고 TCAP, ACTN2, MYH7, VCL등의 순서로 높게 나온다. 근접중심성은 ACTC1이 제일 높게 나오고 TCAP, ACTN2, LDB3, MYH7등의 순서로 높게 나온다. 중개중심성은 ACTC1이 제일 높게 나오고 ACTN2, TCAP, MYH7의 순서로 높게 나온다. Network6의 중심성 분석 결과는 <Table 3.14>을 통해 확인 할 수 있다. 이를 종합해 보면 ACTC1, ACTN2, TCAP, ANKRD1, LDB3, MYH7, MYOZ2, VCL의 순서로 중심도가 높게 나온다. 이렇게 얻은 단백질 네트워크의 중심도 수치를 유전자 네트워크의 중심도 수치에 더하여 protein interaction으로 보정된 유전자 네트워크의 중심도 데이터를 구한다.

Table 3.13 Two-mode network centrality analysis of network5 about germline variant of dilated cardiomyopathy

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Centrality	Centrality	Gene
ACTC1	0.244	ACTC1	0.169	ACTC1	0.168
DMD	0.146	TCAP	0.155	DMD	0.082
TCAP	0.146	DES	0.150	TCAP	0.054
ACTN2	0.098	LDB3	0.150	BAG3	0.044
BAG3	0.098	VCL	0.150	SCN5A	0.044
DES	0.098	DMD	0.148	DES	0.030
FKTN	0.098	LMNA	0.140	LDB3	0.030
LDB3	0.098	ACTN2	0.138	VCL	0.030
LMNA	0.098	FKTN	0.138	LMNA	0.028
MYH7	0.098	MYH7	0.138	SGCD	0.027
SCN5A	0.098	SCN5A	0.138	ACTN2	0.019
SGCD	0.098	CSRP3	0.132	FKTN	0.010
VCL	0.098	MYBPC3	0.132	MYH7	0.010
CAV3	0.049	PLN	0.132		
CSRP3	0.049	RYR2	0.132		
DSP	0.049	TBX20	0.132		
EMD	0.049	TNNC1	0.132		
MYBPC3	0.049	TNNI3	0.132		
PKD1	0.049	TNNT2	0.132		
PLN	0.049	TPM1	0.132		
PSEN1	0.049	TTN	0.132		
PSEN2	0.049	SGCD	0.119		
RYR2	0.049	BAG3	0.114		
TAZ	0.049	PSEN1	0.108		
TBX20	0.049	PSEN2	0.108		
TNNC1	0.049	TAZ	0.108		
TNNI3	0.049	EMD	0.089		
TNNT2	0.049	CAV3	0.088		
TPM1	0.049	DSP	0.008		
TTN	0.049	PKD1	0.008		

The centrality values are the normalized values and the genes are listed in descending order of centrality.

Table 3.14 Two-mode network centrality analysis of network6 about germline variant of hypertrophic cardiomyopathy

Degree centrality		Closeness centrality		Betweenness centrality	
Gene	Centrality	Gene	Gene	Centrality	Gene
ACTC1	0.385	ACTC1	0.146	ACTC1	0.2122
TCAP	0.231	TCAP	0.132	ACTN2	0.0708
ACTN2	0.154	ACTN2	0.125	TCAP	0.0486
LDB3	0.154	LDB3	0.125	MYH7	0.0287
MYH7	0.154	MYH7	0.125	LDB3	0.0091
MYOZ2	0.154	MYOZ2	0.125	MYOZ2	0.0091
VCL	0.154	VCL	0.125	VCL	0.0091
CAV3	0.077	CSRP3	0.119		
CSRP3	0.077	MYBPC3	0.119		
MYBPC3	0.077	MYL2	0.119		
MYL2	0.077	MYL3	0.119		
MYL3	0.077	PLN	0.119		
PLN	0.077	PRKAG2	0.119		
PRKAG2	0.077	TNNC1	0.119		
TNNC1	0.077	TNNI3	0.119		
TNNI3	0.077	TNNT2	0.119		
TNNT2	0.077	TPM1	0.119		
TPM1	0.077	TTN	0.119		
TTN	0.077	CAV3	0.010		

The centrality values are the normalized values and the genes are listed in descending order of centrality.

3.5 유전자 선별

선천적인 확장성 및 비대성 심근증의 원인이 되는 유전자들 중 심근증에서 심부전의 예후를 보이는데 의미가 있는 유전자를 선별하기 위해 로지스틱 회귀분석을 실시한다. 연속형인 독립변수는 위에서 얻은 보정된 중심성 데이터로 설정하고, 범주형인 종속변수는 해당 유전자에 의한 심부전의 유무로 설정을 한다. 심부전의 유무는 유전질환에 대한 데이터 베이스인 Online Mendelian Inheritance in Man (<https://www.omim.org/>, OMIM)에서 유전자의 ‘Clinical Synopsis’에 심부전을 포함하고 있거나 심부전의 증례가 있을 경우 심부전과 관련이 있다고 판단한다. 로지스틱 회귀분석에 사용된 데이터는 <Table 3.15>에서 확인 할 수 있다. 여기서 유전자 네트워크와 단백질 네트워크 모두에서 연결선을 형성하고 있지 않은 ABCC9는 분석에서 제외 한다. 그 뒤 R 프로그래밍 언어를 통해 로지스틱 회귀분석을 실시하면 확장성 심근증에서는 p-value가 8.47×10^{-4} 이고 비대성 심근증에서는 1.95×10^{-2} 로 통계적 유의성이 있다고 나왔다. 본 회귀분석의 회귀모델은 <Figure 3.13>에서 확인 할 수 있다.

다음으로는 receiver operating characteristic (ROC)커브를 그려서 회귀 모델을 평가 하였다. ‘RNA-Seq identifies novel myocardial gene expression signatures of heart failure (Liu et al., 2015)’의 연구 결과 중 심부전에서 발현량이 변하는 129개의 유전자에 대한 확장성 및 비대성 심근증의 원인이 되는 유전자들의 중심성 분석을 수행하고 단백질간의 interaction을 보정한 데이터를 독립변수로, OMIM데이터 베이스의 임상적 결과를 종속변수로 하는 데이터 세트를 앞에서 만든 회귀분석 모델에 적용하여 ROC커브를 그리고 area under the curve (AUC)를 알아보았다. AUC는 확장성 심근증에서 0.840, 비대성 심근증에서 0.787로 나왔으며, ROC 커브는 <Figure 3.14>에서 확인 할 수

있다.

위의 로지스틱 회귀 분석에서 심부전의 위험도를 나타내는 적합값은 0.5 미만의 경우에는 0으로, 0.5이상인 경우에는 1로 반올림을 하여 심부전의 예후를 나타낼 것을 예측하는데, 이를 통해 얻어진 유전자들은, 확장성 심근증 에서는 ACTC1, ACTN2, BAG3, DES, DMD, DSG2, DSP, FKTN, ILK, LAMP2, LDB3, LMNA, MYH7, MYPN, PKD1, PKP2, PLN, PSEN1, RBM20, RYR2, SCN5A, SGCD, TCAP, TPM1, TTN, VCL이고, 비대성 심근증 에서는 ACTC1, ACTN2, ANKRD1, LAMP2, MYOZ2, MYPN, PLN, PRKAG2, TTN, VCL이다. 위 유전자들의 기능은 <Table 3.16>과 <Table 3.17>을 통해 확인 할 수 있다.

위에서 선별된 유전자 외에 중심성은 낮지만 OMIM 데이터베이스에서는 심부전과 관련이 있다고 나오는 유전자들은 확장성 심근증 에서는 MYBPC3, TNNT2, TNNC1, TNNI3가, 비대성 심근증 에서는 MYH7, TPM1, MYBPC3, MYL3가 있다. 위 유전자들의 기능은 <Table 3.18>을 통해 확인 할 수 있다.

Table 3.15 The sum of each centrality indicator and OMIM database search results

a) Centrality and OMIM database result of DCM for comparison4

Gene	Total centrality (%) (gene network)	Total centrality (%) (protein network)	Adjusted centrality (%)	OMIM (heart failure)
ABCC9	0.000	0.000	0.000	Yes
ACTC1	1.255	14.717	7.986	Yes
ACTN2	1.831	3.745	2.788	Yes
ANKRD1	1.208	0.000	0.604	No
BAG3	2.005	4.953	3.479	Yes
CAV3	1.452	1.489	1.470	No
CRYAB	2.080	0.000	1.040	No
CSRP3	0.777	1.891	1.334	No
DES	0.975	4.478	2.726	Yes
DMD	2.976	8.165	5.570	Yes
DSG2	3.514	0.000	1.757	No
DSP	3.514	0.771	2.143	Yes
EMD	1.456	1.499	1.478	No
EYA4	1.287	0.000	0.643	No
FHL2	1.246	0.000	0.623	No
FKRP	1.019	0.000	0.509	No
FKTN	3.518	3.217	3.368	Yes
GATAD1	1.754	0.000	0.877	Yes
ILK	5.286	0.000	2.643	No
LAMA4	0.359	0.000	0.180	No
LAMP2	4.209	0.000	2.105	Yes
LDB3	0.731	4.478	2.604	Yes
LMNA	1.456	4.267	2.862	Yes
MYBPC3	1.000	1.891	1.446	Yes
MYH7	0.731	3.217	1.974	Yes
MYPN	4.393	0.000	2.197	Yes
NEXN	3.267	0.000	1.633	No
PDLIM3	0.777	0.000	0.389	No

a) Centrality and OMIM database result of DCM for comparison4 (cont.)

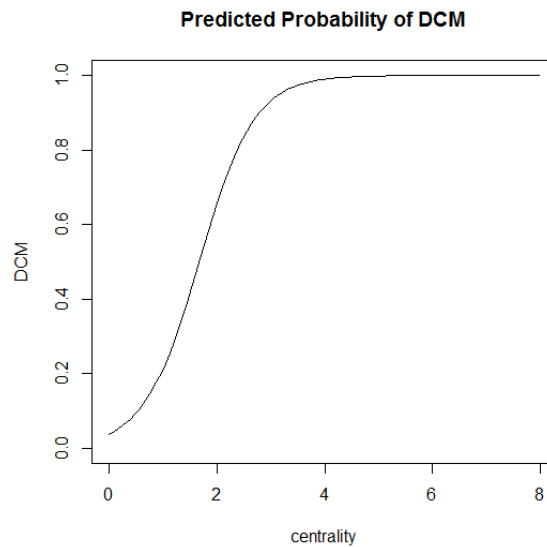
Gene	Total centrality (%) (gene network)	Total centrality (%) (protein network)	Adjusted centrality (%)	OMIM (heart failure)
PKD1	3.521	0.771	2.146	No
PKP2	3.449	0.000	1.725	No
PLN	3.079	1.891	2.485	Yes
PRDM16	0.000	0.000	0.000	No
PSEN1	5.388	1.670	3.529	Yes
PSEN2	1.597	1.670	1.633	Yes
RBM20	4.643	0.000	2.321	Yes
RYR2	2.424	1.891	2.157	No
SCN5A	2.386	5.174	3.780	Yes
SGCD	3.038	4.041	3.540	Yes
TAZ	0.000	1.670	0.835	No
TBX20	1.338	1.891	1.614	No
TCAP	0.748	6.620	3.684	Yes
TMPO	1.281	0.000	0.640	No
TNNC1	1.255	1.891	1.573	Yes
TNNI3	0.000	1.891	0.946	Yes
TNNT2	0.748	1.891	1.319	Yes
TPM1	1.831	1.891	1.861	Yes
TTN	2.078	1.891	1.985	Yes
TXNRD2	2.051	0.000	1.026	No
VCL	5.068	4.478	4.773	Yes

b) Centrality and OMIM database result of HCM for comparison4

Gene	Total centrality (%) (gene network)	Total centrality (%) (protein network)	Adjusted centrality (%)	OMIM (heart failure)
ACTC1	0.000	25.994	12.997	Yes
ACTN2	4.632	10.190	7.411	Yes
ANKRD1	3.922	4.180	4.051	No
CALR3	0.000	2.469	1.235	No
CAV3	3.077	2.051	2.564	No
CSRP3	2.520	3.676	3.098	No
FHL1	5.418	0.785	3.102	No
GLA	0.000	0.000	0.000	No
JPH2	0.000	0.000	0.000	No
LAMP2	9.272	0.000	4.636	Yes
LDB3	2.339	4.106	3.223	No
MYBPC3	2.339	2.891	2.615	Yes
MYH7	1.795	4.106	2.951	Yes
MYL2	0.000	2.891	1.446	No
MYL3	0.000	2.891	1.446	Yes
MYLK2	0.000	0.000	0.000	No
MYOZ2	9.807	4.106	6.956	Yes
MYPN	9.011	0.000	4.506	Yes
NEXN	6.274	0.000	3.137	No
PLN	13.743	2.891	8.317	Yes
PRKAG2	6.871	2.891	4.881	No
TCAP	0.000	5.320	2.660	Yes
TNNC1	0.000	2.891	1.446	No
TNNI3	0.000	2.891	1.446	No
TNNT2	2.339	2.891	2.615	No
TPM1	0.000	2.891	1.446	Yes
TTN	6.010	2.891	4.451	Yes
VCL	10.630	4.106	7.368	Yes

Total centrality is the sum of all the centrality factors, followed by the percentage of the total centrality, and the adjusted centrality is the mean of these. If the OMIM database reports on the onset of heart failure, the OMIM result is 'Yes'. Otherwise, the OMIM result is 'No'.

a)



b)

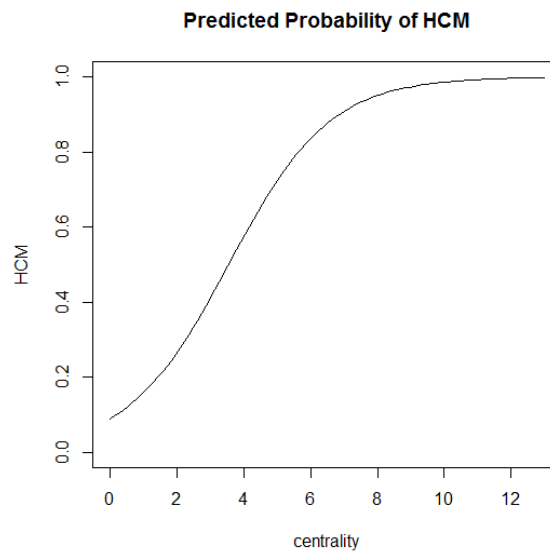
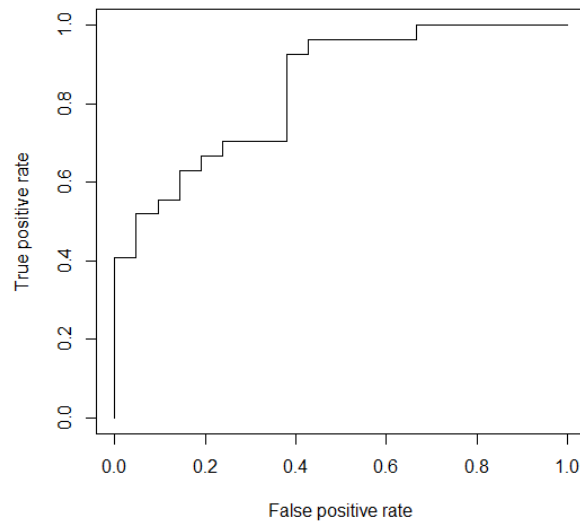


Figure 3.13 Logistic regression curve for gene selection. a) The independent variable of the regression curve is the centrality of the causative gene of DCM and the dependent variable is the OMIM result. $p\text{-value} = 8.47 \times 10^{-4}$. b) The independent variable of the regression curve is the centrality of the causative gene of HCM and the dependent variable is the OMIM result. $p\text{-value} = 1.95 \times 10^{-2}$.

a)



b)

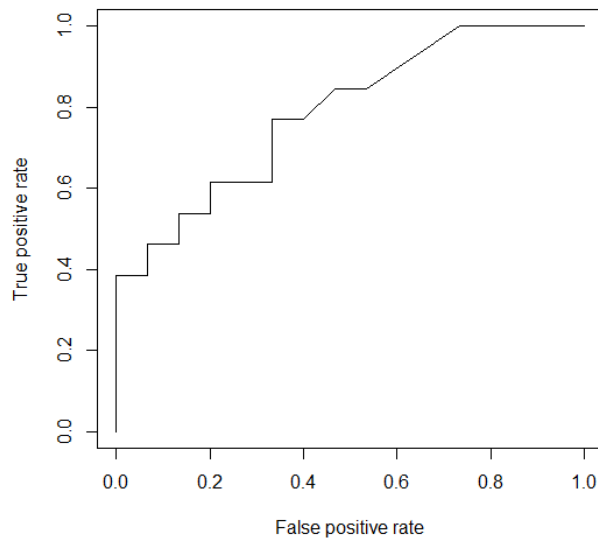


Figure 3.14 ROC curve for logistic regression curve. Centrality analysis is performed between the genes whose expression levels change in heart failure in other study and the genes responsible for DCM and HCM. The result is applied to the regression model to draw the ROC curve and calculate AUC. a) DCM, AUC = 0.840. b) HCM, AUC = 0.787.

Table 3.16 Genes that can predict heart failure with DCM prognosis

Gene	Protein	Function
ACTC1	Actin, alpha cardiac muscle 1	Involved in various types of cell motility and ubiquitously expressed in all eukaryotic cells
ACTN2	Alpha-actinin-2	F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures
BAG3	BAG family molecular chaperone regulator 3	Promoting the release of ADP from the HSP70 and HSC70 proteins thereby triggering client/substrate protein release
DES	Desmin	Sarcomeric microtubule-anchoring protein
DMD	Dystrophin	Anchors the extracellular matrix to the cytoskeleton via F-actin
DSG2	Desmoglein-2	Component of intercellular desmosome junctions
DSP	Desmoplakin	Major high molecular weight protein of desmosomes
FKTN	Fukutin	Interact with and reinforce a large complex encompassing the outside and inside of muscle membranes
ILK	Integrin-linked protein kinase	Receptor-proximal protein kinase regulating integrin-mediated signal transduction
LAMP2	Lysosome-associated membrane glycoprotein 2	Plays an important role in chaperone-mediated autophagy
LDB3	LIM domain-binding protein 3	Adapter in striated muscle to couple protein kinase C-mediated signaling via its LIM domains to the cytoskeleton
LMNA	Prelamin-A/C	Accelerate smooth muscle cell senescence

Table 3.16 Genes that can predict heart failure with DCM prognosis (cont.)

Gene	Protein	Function
MYH7	Myosin-7	Muscle contraction
MYPN	Myopalladin	Component of the sarcomere that tethers together nebulin and nebulin to alpha-actinin
PKD1	Polycystin-1	Involved in renal tubulogenesis
PKP2	Plakophilin-2	May play a role in junctional plaques
PLN	Cardiac phospholamban	Reversibly inhibits the activity of ATP2A2 in cardiac sarcoplasmic reticulum
PSEN1	Presenilin-1	Catalytic subunit of the gamma-secretase complex
RBM20	RNA-binding protein 20	Regulates splicing of TTN
RYR2	Ryanodine receptor 2	Formation of heterotetramers with RYR3 and modulate calcium channel activity
SCN5A	Sodium channel protein type 5 subunit alpha	Mediates the voltage-dependent sodium ion permeability of excitable membranes
SGCD	Delta-sarcoglycan	Forms a link between the F-actin cytoskeleton and the extracellular matrix
TCAP	Telethonin	Mediates the antiparallel assembly of titin (TTN) molecules at the sarcomeric Z-disk
TPM1	Tropomyosin alpha-1 chain	Calcium dependent regulation of vertebrate striated muscle contraction
TTN	Titin	Key component in the assembly and functioning of vertebrate striated muscles
VCL	Vinculin	Actin filament binding protein involved in cell-matrix adhesion and cell-cell adhesion

The fitting value of above genes are more than 0.5 in the logistic regression analysis. It is possible to predict the prognosis of heart failure if the DCM is caused by the above gene mutation.

Table 3.17 Genes that can predict heart failure with HCM prognosis

Gene	Protein	Function
ACTC1	Actin, alpha cardiac muscle 1	Involved in various types of cell motility and ubiquitously expressed in all eukaryotic cells
ACTN2	Alpha-actinin-2	F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures
ANKRD1	Ankyrin repeat domain-containing protein 1	Nuclear transcription factor that negatively regulates the expression of cardiac genes
LAMP2	Lysosome-associated membrane glycoprotein 2	Plays an important role in chaperone-mediated autophagy
MYOZ2	Myozenin-2	Serve as intracellular binding proteins involved in linking Z line proteins
MYPN	Myopalladin	Component of the sarcomere that tethers together nebulin and nebulin to alpha-actinin
PLN	Cardiac phospholamban	Reversibly inhibits the activity of ATP2A2 in cardiac sarcoplasmic reticulum
PRKAG2	5'-AMP-activated protein kinase subunit gamma-2	Regulating cellular energy metabolism
TTN	Titin	Key component in the assembly and functioning of vertebrate striated muscles
VCL	Vinculin	Actin filament binding protein involved in cell-matrix adhesion and cell-cell adhesion

The fitting value of above genes are more than 0.5 in the logistic regression analysis. It is possible to predict the prognosis of heart failure if the HCM is caused by the above gene mutation.

Table 3.18 Genes that can predict heart failure with low adjusted centrality

Gene	Protein	Function
< Genes associated with dilated cardiomyopathy >		
MYBPC3	Myosin-binding protein C, cardiac-type	Modulate muscle contraction or play a more structural role
TNNC1	Troponin C, slow skeletal and cardiac muscles	Binding of calcium to Tn-C abolishes the inhibitory action of Tn on actin filaments
TNNI3	Troponin I, cardiac muscle	Confers calcium-sensitivity to striated muscle actomyosin ATPase activity
TNNT2	Troponin T, cardiac muscle	Confers calcium-sensitivity to striated muscle actomyosin ATPase activity.
< Genes associated with hypertrophic cardiomyopathy >		
MYBPC3	Myosin-binding protein C, cardiac-type	Modulate muscle contraction or play a more structural role
MYH7	Myosin-7	Muscle contraction
MYL3	Myosin light chain 3	Regulatory light chain of myosin
TPM1	Tropomyosin alpha-1 chain	Calcium dependent regulation of vertebrate striated muscle contraction

The fitting value of above genes are less than 0.5 in the logistic regression analysis, but there are cases of heart failure in OMIM.

제 4 장. 고 찰

4.1 심근증으로 인한 심부전의 마커유전자 탐색

심부전은 말기의 경우 사망률이 높고 고비용의 치료가 필요한 만큼 조기에 진단을 하고 적절한 치료를 하는 것이 심부전 환자의 수명을 연장 시키고 삶의 질을 향상 시키는데 중요하다고 할 수 있다. 유전적인 원인에 의한 심부전의 경우 많은 경우가 심근증으로부터 비롯된다. 그렇기 때문에 심근증 환자들의 예후로 심부전이 나타날 것을 예측 하는 것은 적절한 치료와 모니터링을 통해 질환이 심부전으로 악화 되는 것을 줄이는데 도움이 될 수 있을 것이다. 또한 의료 기술의 발전으로 Clustered regularly interspaced short palindromic repeats (CRISPR)같이 유전자에 대한 직접적인 처치가 가능해 지고 있기 때문에 질병에 대한 분자수준의 연구는 질병의 발병 메커니즘을 이해하는데 뿐만 아니라 분자수준의 치료를 하는데 많은 정보를 제공 할 수 있을 것이다(Zhang et al., 2014).

Comparison4의 1부 네트워크 분석에서는 C1QTNF7, ECM2, FAM188A, HNMT, ITIH5, NAMPT, OIP5-AS1, PTN, SMOC2, SULF1의 중심성이 높게 나왔는데, 이는 심근증에서 심부전의 예후가 나타나게 하는데 중요한 역할을 하는 유전자들이라 볼 수 있다. 이는 심근증이 심부전으로 악화되는 과정에 대한 연구 및 치료의 타겟이 될 수 있다. 확장성 및 비대성 심근증의 원인 유전자와 Comparison4간의 2부 네트워크 분석에서 확장성 심근증의 경우 ACTC1, ACTN2, BAG3, DES, DMD, DSG2, DSP, FKTN, ILK, LAMP2, LDB3, LMNA, MYH7, MYPN, PKD1, PKP2, PLN, PSEN1, RBM20, RYR2, SCN5A,

SGCD, TCAP, TPM1, TTN, VCL의 중심성이, 비대성 심근증의 경우에는 ACTC1, ACTN2, ANKRD1, LAMP2, MYOZ2, MYPN, PLN, PRKAG2, TTN, VCL의 중심성이 각각 높게 나왔다. 이는 위의 유전자의 선천적인 변이로 인한 심근증의 경우 예후로 심부전이 나타날 가능성이 높다는 것을 유추 할 수 있으며, 유전적인 요인으로 인한 심근증의 경우 심부전의 예후를 예측 하는데 위의 유전자들을 바이오마커로 활용 할 수 있을 가능성을 제시해 준다.

위에서 선별된 유전자 외에 중심성은 낮지만 OMIM 데이터베이스에서는 심부전과 관련이 있다고 나오는 유전자들은 확장성 심근증 에서는 MYBPC3, TNNC1, TNNI3, TNNT2가, 비대성 심근증 에서는 MYBPC3, MYH7, MYL3, TPM1등이 있는데 위의 유전자들은 Kyoto Encyclopedia of Genes and Genomes (KEGG) 데이터베이스의 Cardiac muscle contraction pathway에서 Calcium signaling pathway와 myosin 사이에 위치하거나 myosin과 직접 연결되는 단백질을 코딩하고 있는 유전자들이다. 이는 유전자 발현상관성과 단백질간의 interaction이 낮아도 심장의 수축/이완 과정에 중요한 역할을 하는 유전자의 변이는 심부전을 불러 올 수 있다는 것을 유추 할 수 있다.

그 외에 확장성 심근증에서 ABCC9은 유전자 네트워크와 단백질 네트워크 모두에서 연결선을 형성하지 않았지만, ABCC9유전자의 frameshift (FS1524)와 missense (A1513T)의 변이에 의해 심부전 증상이 나타난 경우도 있다. ABCC9은 cardiac ATP-sensitive potassium channel의 SUR2A 서브유닛을 코딩 하고 있는데, 보존서열상에 위치하고 있는 위의 두 변이로 인해 cardiac ATP-sensitive potassium channel 이 제 기능을 하지 못하여 DCM이 발생하고 심부전으로 증세가 악화 된다고 알려져 있다(Bienengraeber et al., 2004).

4.2 활용 방안

본 연구의 목적은 심근증에 대한 유전자, 단백질체, 임상적 정보를 통합한 데이터 베이스를 구축하여 심근증의 연구 및 임상에서 이용 될 수 있는 정보를 제공하고, 유전자 발현 수준에 대한 연구 결과들을 통상적인 진단검사에 적용 할 수 있는 방법을 연구 하는데 있다. 먼저 심근증에 대한 유전학적 정보를 효율적으로 얻을 수 있도록 하기 위해 심근증과 관련이 있는 유전자의 서열 및 해당 유전자가 암호화 하고 있는 단백질의 아미노산 서열과 단백질의 구조, 유전자의 변이가 가지고 있는 임상적 의의에 대한 정보들을 이용하여 데이터베이스를 구축 하였다. 데이터베이스의 검색시스템을 이용하여 clinical significance가 ‘pathogenic’이나 ‘likely pathogenic’인 유전자들을 case excess의 순서대로 나열하면 다음과 같다. DCM에서는 TTN(0.13689), RBM20(0.04355), DSP(0.04346), MYH7(0.03849), LMNA(0.03782), LAMA4(0.02677), TNNT2(0.02629), NEXN(0.02449), TNNC1(0.01863), TPM1(0.01804), VCL(0.01187), TCAP(0.00944), DES(0.00535), MYBPC3(0.00528), ABCC9(0.00482), TNNI3(0.00418), ACTC1(0.00389), LDB3(0.00364), SGCD(0.0034), PLN(0.00317), CRYAB(0.00263), MYH6(0.00013)의 순서로 case excess가 높게 나오고, HCM에서는 MYBPC3(0.17062), MYH7(0.1282), TNNI3(0.02005), TNNT2(0.01708), TPM1(0.01398), MYL2(0.00923), FHL1(0.00789), PRKAG2(0.00525), NEXN(0.0051), ACTC1(0.00462), LAMP2(0.0044), PLN(0.00275), CSRP3(0.00184)의 순서로 case excess가 높게 나온다. 실험군의 유전자 변이 빈도에서 대조군의 변이 빈도를 빼는 것으로 나타내는 case excess는 수치가 높을수록 실험군 에서 변이가 많이 발생한다고 해석 할 수 있다. 이는 단순히 실험군의 변이 빈도를 보는 것보다 유전자에 대한 유병률을 더 자세하게 나타낼 수 있는 지표가 될 수 있을 것이다. 예를 들어 MYBPC3의 case excess는 DCM

에서는 0.00528, HCM에서는 0.17062로 나타나는데, 이는 만약 MYBPC3에 변이가 있을 경우 DCM보다는 HCM이 발병할 확률이 높다는 것을 예측 할 수 있다. 또한 non-truncating variants와 truncating variants의 case excess를 각각 볼 수 있는데, 이는 유전자에 발생하는 변이의 종류에 따라 질병이 발병 할 확률을 예측 할 수 있게 해준다. 예를 들어 HCM에서 MYH7의 경우 non-truncating variants의 case excess는 0.12786을 truncating variants의 case excess는 0.00034을 각각 나타내는데 이는 MYH7에 truncating variant인 frameshift변이가 발생 했을 때보다 non-truncating variant인 missense (clinical significance가 pathogenic 혹은 likely pathogenic)변이가 발생 했을 때 HCM이 발병할 확률이 더 높다는 것을 예측 할 수 있게 해 준다. 따라서 본 데이터 베이스를 기초로 한 유전학적 정보들은 분자 진단 및 의학적 연구, 인류의 건강 증진을 목적으로 하는 보건학적 연구 등에 활용 될 수 있을 것이다.

또한 유전자 발현 수준에 대한 연구 결과들을 진단검사에 적용 할 수 있는 방법에 대한 연구는 마이크로어레이 데이터를 이용한 차별발현 유전자들의 탐색과 발현상관성에 기초한 유전자 네트워크 분석을 통해 이루어 졌다. 이를 통해 특정 조직에서 일어나는 유전자들의 발현양상의 변화와 상관성이 있는 질병관련 유전자들을 탐색하여 실제 진단 및 질병의 예후를 예측하는데 사용 될 수 있는 마커유전자들을 찾아 말초혈액을 이용한 진단검사에 활용 될 수 있도록 하였다. 아울러 차별발현 유전자들을 이용한 유전자 네트워크 구성과 이에 대한 분석의 결과로 특정 표현형이 나타나게 하는 유전자들의 상관관계 및 중요도가 높은 유전자들을 찾을 수 있었다. 이는 특정 표현형에 대한 생물학적 경로 및 질병의 발생 기전을 연구하는데 활용 할 수 있을 뿐만 아니라 치료약의 개발 및 유전자 치료의 대상 유전자들을 선별하는데도 이용 할 수 있을 것이다.

4.3 보건학적 연구에 적용

본 연구에서 구축한 데이터베이스는 심근증과 관련이 있는 유전자의 서열 및 변이, 이 변이가 가지는 임상적 의의, 해당 유전자가 발현하는 단백질의 아미노산 서열과 단백질의 구조에 대한 정보를 통합하였다. 심근증은 가족력이 높은 질환으로, 다른 심혈관계 질환들은 나이가 많을수록 발병의 위험이 높아지지만 심근증은 다양한 연령대에서 발병 하고 특히 아동기 돌연사의 주요한 원인 중 에 하나이다. 유전자 검사를 통한 심근증의 조기 진단은 적절한 치료를 받을 수 있는 기회를 제공하는데 중요하며 특히 본인의 증상을 표현하기 어려운 아동들의 조기 진단에 큰 가치가 있다. 따라서 본 데이터베이스의 유전학적 정보들을 이용함으로써 심근증의 조기 진단의 정확성을 높일 수 있을 것이다.

또한 유전자 네트워크분석의 결과들을 통해 심근증의 특정 예후와 관련이 있는 유전자 들을 유추 할 수 있는데 이는 예후의 예측을 위한 마커유전자를 찾는 데 도움을 줄 수 있다. 또한 심장에서 일어나는 유전자 발현 정도의 차이와 유전자들 사이의 발현 상관성 및 중요도에 대한 분석 결과들은 질환의 기전과 경로에 대한 연구에 도움을 줄 수 있으며, 이로 인하여 신약의 개발과 치료법이 발전하는데 기여 할 수 있을 것이다.

위와 같이 본 연구에서의 심근증에 대한 유전학적 데이터베이스 구축과 유전자 네트워크 분석 결과들은 심근증의 발병 기전에 대한 이해뿐만 아니라 심근증의 조기진단과 예후를 예측 할 수 있게하여 적절한 치료와 지속적인 모니터링이 가능하게 해 줄 것이다. 이로 인해 심근증 환자들의 수명의 연장과 삶의 질이 향상 되는 것 뿐만 아니라 관련 의료비용의 절감도 가능케 하여 보건학적으로 높은 성과를 기대 할 수 있다.

제 5 장. 결 론

질병의 치료를 위한 분자적 기전을 밝히고 유전자 변이와 표현형의 관계를 밝히는 한편 연구 결과를 임상치료에 적용하기 위한 생물학적 연구에는 인간의 질병에 대한 정보가 꼭 필요하다. 휴먼게놈프로젝트 이후 NGS등의 염기서열분석법들의 발달로 인간의 유전학적 정보들이 증가하고 있는데 이러한 정보들을 효율적으로 분석하고 이용하기 위해서는 데이터베이스의 구축이 중요하다. 유전성 심근증은 유전학적으로 다양한 형태의 변이가 나타나기 때문에 유전학적 검사를 통해 심근증의 조기 진단 및 예후의 예측, 치료의 단서를 제공 받을 수 있다. 심근증에 대한 유전학적인 정보를 효율적으로 이용하기 위해 본 연구에서는 여러 데이터베이스에 산재해 있는 심근증의 유전학적인 정보들을 수집하여 통합적인 데이터베이스를 구축하였다. 이를 통해 유전자 및 단백질에 대한 정보를 얻을 수 있도록 하였는데 특히 유전학적 정보를 진단에 활용 할 수 있도록 유전자의 변이가 임상적으로 질환을 일으키는데 영향을 미치는 정도에 대한 정보를 포함 시켰다. 단백질의 구조 변화에 영향이 큰 삽입/결실변이나 넌센스변이는 대부분 임상적으로 질병을 일으키는데 영향이 크다고 알려졌는데, 그 외에 단백질의 구조변화를 예측하기 힘든 미스센스변이에 대한 임상적 의의를 이를 통해 확인 할 수 있다. 최근 NGS등을 통해 단시간에 여러 유전자에 대한 변이 유무의 검사가 가능해 지고 있기 때문에 이러한 유전학적 정보는 신속하고 정확한 심근증의 진단에 크게 기여 할 수 있을 것이다.

심근증으로 인한 심부전 환자의 마이크로어레이 데이터를 통하여 확장성 및 비대성 심근증에서 심부전의 예후를 나타나게 하는

유전자들을 유추 할 수 있다. 발현 상관성 정보를 이용하여 유전자 네트워크를 구성하면 전체 네트워크가 두 개의 서브 네트워크로 나뉘며 하나의 네트워크는 up-regulated되는 유전자들로 구성이 되어있고 다른 하나는 up-regulated되는 유전자들과 down-regulated되는 유전자들로 함께 구성이 되어있다. 이렇게 구성된 네트워크의 중심성 분석을 통해 네트워크에서 허브 등의 중요한 역할을 하는 위치에 있는 유전자들을 유추 할 수 있는데 C1QTNF7, ECM2, FAM188A, HNM1, ITIH5, NAMPT, OIP5-AS1, PTN, SMOC2, SULF1이 그러한 역할을 하는 유전자들이다. 위의 유전자들 모두 두 서브네트워크가 인접한 곳에 위치하여 다른 서브네트워크를 구성하는 유전자들과 관계를 형성하며 네트워크의 발현량 조절에 중요한 역할을 한다는 것을 알 수 있다. 따라서 다른 유전자들 보다 위의 유전자들에 변이가 있을 경우 임상적으로 더 심각한 결과를 초래 할 수 있다는 것을 유추 할 수 있는 한편, 치료약 및 치료의 표적으로 삼을 수 있을 것이다.

다음으로 서로 성질이 다른 개체간의 관계를 알 수 있는 2부 네트워크 분석을 통해 심부전의 예후가 나타나게 하는 유전자들과 선천적인 변이가 심근증의 원인이 되는 유전자들간의 관계를 알 수 있는데 이를 통해 선천적인 유전자 변이에 의한 심근증 환자들 중심부전의 예후를 보일 것 이라고 예측 할 수 있는 마커유전자들을 유추 할 수 있다. 심부전의 예후가 나타나게 하는 유전자들을 중심으로 2부네트워크 분석을 수행하면 확장성 심근증에서는 ARMC8, EHD4, FLNC, MYH6, NME7, OIP5-AS1, PYROXD1, SCN2B, SSPN이, 비대성 심근증에서는 ARMC8, FLNC, MYO16, PYROXD1, SCN2B, SSPN이 심근증의 원인이 되는 유전자들과 연관성이 크다는 것을 알 수 있다. 위의 유전자들은 모두 up-regulated되는 유전자들과 down-regulated되는 유전자들이 이루는 서브 네트워크에 속하는데 이는 유

전적인 원인으로 심근증이 발병할 경우 심부전의 예후를 나타내게 하는 데는 위의 서브네트워크에 속해있는 유전자들이 큰 역할을 한다는 것을 알 수 있다. 선천적인 변이가 심근증의 원인이 되는 유전자들을 중심으로 한 2부네트워크 분석에서는, 확장성 심근증의 경우 ACTC1, ACTN2, BAG3, DES, DMD, DSG2, DSP, FKTN, ILK, LAMP2, LDB3, LMNA, MYH7, MYPN, PKD1, PKP2, PLN, PSEN1, RBM20, RYR2, SCN5A, SGCD, TCAP, TPM1, TTN, VCL의 중심성이, 비대성 심근증의 경우에는 ACTC1, ACTN2, ANKRD1, LAMP2, MYOZ2, MYPN, PLN, PRKAG2, TTN, VCL의 중심성이 높게 나왔는데, 이를 통해 위의 유전자들이 심부전의 예후가 나타나게 하는 유전자들과 연관성이 크다는 것을 알 수 있다. 이는 위의 유전자의 변이로 인한 심근증이 유발 될 경우 심부전의 예후를 보일 가능성이 크다는 것을 유추 할 수 있다. 즉, 위의 심부전의 예후가 나타나게 하는 유전자들과 연관성이 큰 유전자들 및 심부전의 예후가 나타나게 하는 유전자들의 네트워크에 속해있는 MYH6는 각각의 심근증 환자들에게서 해당 유전자에 변이가 있을 경우 심부전의 예후가 나타날 가능성이 높다는 것을 예측 할 수 있게 해주는 마커유전자로 활용 될 수 있을 것이다. 이러한 마커유전자들의 변이는 말초혈액을 이용한 통상적인 진단검사를 통해 검출 할 수 있고, 그 결과로 특정 조직에서의 유전자 발현양상의 변화로 인한 질병의 진단 및 예후를 예측 할 수 있게 해 주는데, 이는 병세의 지속적인 모니터링과 치료의 방향을 결정 할 수 있게 해주어 환자의 수명연장과 삶의 질 향상에 크게 기여 할 수 있을 것이다.

참 고 문 헌

- Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's online Mendelian inheritance in man (OMIM). *Nucleic acids research*. 37(suppl 1):D793-D796 (2009).
- Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM). *Human mutation*. 32(5):564-567 (2011).
- Askoxylakis V, Thieke C, Pleger ST, Most P, Tanner J, Lindel K, Katus HA, Debus J, Bischof, M. Long-term survival of cancer patients compared to heart failure and stroke: a systematic review. *BMC cancer*. 10(1):105 (2010).
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic acids research*. 41(D1):D36-D42 (2013).
- Bharucha T, Lee KJ, Daubeney PE, Nugent AW, Turner C, Sholler GF, Robertson T, Robert J, Ramsay J, Carlin JB, Colan SD. Sudden death in childhood cardiomyopathy: results from a long-term national population-based study. *Journal of the American College of Cardiology*. 65(21):2302-2310 (2015).
- Bienengraeber M, Olson TM, Selivanov VA, Kathmann EC, O'Cochlain F, Gao F, Karger AB, Ballew JD, Hodgson DM, Zingman LV, Pang YP, Alekseev AE, Terzic A. ABCC9 mutations identified in human dilated cardiomyopathy disrupt catalytic KATP channel gating. *Nature genetics*. 36(4):382-387 (2004).
- Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 19(2):185-193 (2003).
- Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. *Social networks*. 28(4):466-484 (2006).
- Charron P, Dubourg O, Desnos M, Isnard R, Hagege A, Bonne G, Carrier L, Tesson F, Bouhour JB, Buzzi JC, Feingold J. Genotype-phenotype correlations in familial hypertrophic cardiomyopathy. *European Heart Journal*. 19(1):139-145 (1998).
- Charron P, Arad M, Arbustini E, Basso C, Bilinska Z, Elliott P, Helio T, Keren A, McKenna WJ, Monserrat L, Pankuweit S, Perrot A, Rapezzi C, Ristic A, Seggewiss H, van Langen I, Tavazzi L. Genetic counselling and testing in cardiomyopathies: a position statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases. *European heart journal*. 31(22):2715-2726 (2010).

Chiaretti S, Li X, Gentleman R, Vitale A, Vignetti M, Mandelli F, Foa R. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood* 103(7):2771-2778 (2004).

Deshpande SR, Herman HK, Quigley PC, Shinnick JK, Cundiff CA, Caltharp S, Shehata BM. Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia (ARVC/D): Review of 16 Pediatric Cases and a Proposal of Modified Pediatric Criteria. *Pediatric cardiology*. 1-10 (2016).

DeWitt MM, MacLeod HM, Soliven B, McNally EM. Phospholamban R14 deletion results in late-onset, mild, hereditary dilated cardiomyopathy. *Journal of the American College of Cardiology*. 48(7): 1396-1398 (2006).

Elliott P, Andersson B, Arbustini E, Bilinska Z, Cecchi F, Charron P, Dubourg O, Kuhl U, Maisch B, McKenna WJ, Monserrat L. Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. *European heart journal*(2007).

Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, Charron P, Hagege AA, Lafont A, Limongelli G, Mahrholdt H, McKenna WJ. ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy. *European heart journal*. ehv284(2014).

Frazier A, Judge DP, Schulman SP, Johnson N, Holmes KW, Murphy AM. Familial hypertrophic cardiomyopathy associated with cardiac β -myosin heavy chain and troponin I mutations. *Pediatric cardiology*. 29(4):846-850 (2008).

Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Franco S. Heart disease and stroke statistics-2013 update. *Circulation*, 127(1) (2013).

Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Fullerton HJ. Heart disease and stroke statistics-2014 update. *Circulation*. 129(3) (2014).

Golbus JR, Puckelwartz MJ, Dellefave-Castillo L, Fahrenbach JP, Nelakuditi V, Pesce LL, Pytel P, McNally EM. Targeted analysis of whole genome sequence data to diagnose genetic cardiomyopathy. *Circulation: Cardiovascular Genetics*. 7(6):751-759 (2014).

GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 348(6235):648-660 (2015).

Harrison SM, Riggs ER, Maglott DR, Lee JM, Azzariti DR, Niehaus A, Ramos EM, Martin CL, Landrum MJ, Rehm HL. Using ClinVar as a Resource to Support Variant Interpretation. *Current Protocols in Human Genetics*. 8-16 (2016).

- Harvey PA, Leinwand LA. Cellular mechanisms of cardiomyopathy. *The Journal of cell biology*. 194(3): 355-365 (2011).
- Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, Conner L, DePalma SR, McDonough B, Sparks E, Teodorescu DL. Truncations of titin causing dilated cardiomyopathy. *New England Journal of Medicine*. 366(7):619-628 (2012).
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 4(1):44-57 (2009).
- Kapoor M, Seth S, Rao VR. Clinical genetic aspects of cardiomyopathies. *Journal of the Practice of Cardiovascular Sciences*. 1(2):120 (2015).
- Kensler RW, Shaffer JF, Harris SP. Binding of the N-terminal fragment C0-C2 of cardiac MyBP-C to cardiac F-actin. *Journal of structural biology*. 174(1):44-51 (2011).
- Kramer CM, Appelbaum E, Desai MY, Desvigne-Nickens P, DiMarco JP, Friedrich MG, Geller N, Heckler S, Ho CY, Jerosch-Herold M, Ivey EA. Hypertrophic Cardiomyopathy Registry: The rationale and design of an international, observational study of hypertrophic cardiomyopathy. *American heart journal*. 170(2):223-230 (2015).
- Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, Mungall CJ, Binder JX, Malone J, Vasant D, Parkinson H. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic acids research*. 43(D1):D1071-D1078 (2015).
- Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research*. 42(D1):D980-D985(2014).
- Lazzarini E, Jongbloed JD, Pilichou K, Thiene G, Basso C, Bikker H, Charbon B, Swertz M, van Tintelen JP, der Zwaag PA. The ARVD/C genetic variants database: 2014 update. *Human mutation*. 36(4):403-410 (2015).
- Lange S, Xiang F, Yakovenko A, Vihola A, Hackman P, Rostkova E, Kristensen J, Brandmeier B, Franzen G, Hedberg B, Gunnarsson LG. The kinase domain of titin controls muscle gene expression and protein turnover. *Science*. 308(5728):1599-1603 (2005).
- Lee JH, Lim NK, Cho MC, Park HY. Epidemiology of heart failure in Korea: present and future. *Korean Circulation Journal*. 46(5):658-664 (2016).
- Li W, Chen L, He W, Li W, Qu X, Liang B, Gao Q, Feng C, Jia X, Lv Y, Zhang, S. Prioritizing Disease Candidate Proteins in Cardiomyopathy-Specific Protein-Protein

Interaction Networks Based on “Guilt by Association” Analysis. *PloS one*. 8(8):e71191 (2013).

Limongelli G, Elliott P, Charron P, Mogensen J, McKeown PP. Approaching genetic testing in cardiomyopathies (ESC Council for Cardiology Practice) (2012).

Liu Y, Morley M, Brandimarto J, Hannenhalli S, Hu Y, Ashley EA, Tang WH, Moravec CS, Margulies KB, Cappola TP, Li M. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics*. 105(2):83-89 (2015).

Marcus FI, Edson S, Towbin JA. Genetics of arrhythmogenic right ventricular cardiomyopathy: a practical guide for physicians. *Journal of the American College of Cardiology*. 61(19):1945-1948 (2013).

Maron BJ, Niimura H, Casey SA, Soper MK, Wright GB, Seidman JG, Seidman CE. Development of left ventricular hypertrophy in adults with hypertrophic cardiomyopathy caused by cardiac myosin-binding protein C gene mutations. *Journal of the American College of Cardiology*. 38(2):315-321 (2001).

Maron BJ, Carney KP, Lever HM, Lewis JF, Barac I, Casey SA, Sherrid MV. Relationship of race to sudden cardiac death in competitive athletes with hypertrophic cardiomyopathy. *Journal of the American College of Cardiology*, 41(6):974-980 (2003).

Maron BJ, Maron MS. Hypertrophic cardiomyopathy. *The Lancet*. 381(9862):242-255(2013).

McCartan C, Mason R, Jayasinghe SR, Griffiths LR. Cardiomyopathy classification: ongoing debate in the genomics era. *Biochemistry research international*. 2012 (2012).

McNally EM, Barefield DY, Puckelwartz MJ. The genetic landscape of cardiomyopathy and its role in heart failure. *Cell metabolism* 21(2):174-182 (2015).

McKenna WJ, England D, Doi YL, Deanfield JE, Oakley CM, Goodwin JF. Arrhythmia in hypertrophic cardiomyopathy. I: Influence on prognosis. *British heart journal*. 46(2):168-172 (1981).

McNally EM, Puckelwartz MJ. Genetic variation in cardiomyopathy and cardiovascular disorders. *Circulation journal: official journal of the Japanese Circulation Society*, 79(7): 1409 (2015).

Miranda JO, Costa L, Rodrigues E, Teles EL, Baptista MJ, Areias JC. Paediatric dilated cardiomyopathy: clinical profile and outcome. The experience of a tertiary centre for paediatric cardiology. *Cardiology in the Young*. 25(02):333-337 (2015).

Morita H, Seidman J, Seidman CE. Genetic causes of human heart failure. *The Journal of clinical investigation*. 115(3):518-526 (2005).

- Mulligan MK, Mozhui K, Prins P, Williams RW. GeneNetwork – A toolbox for systems genetics. In Systems Genetics, Methods in Molecular Biology in press (2016).
- Newman M. Networks: an introduction. Oxford university press (2010).
- Ohn JH, Kim J, Kim JH. Social network analysis of gene expression data. In AMIA Annual Symposium Proceedings. American Medical Informatics Association. Vol.2003:p.958 (2003).
- Otte E, Rousseau R. Social network analysis: a powerful strategy, also for the information sciences. Journal of information Science. 28(6):441-453 (2002).
- Pasternak J. An introduction to human molecular genetics: mechanisms of inherited diseases. John Wiley & Sons (2005).
- Pundir S, Martin MJ, O'Donovan, C. UniProt Tools. Current Protocols in Bioinformatics. 1-29 (2002).
- Roberts JM. Correspondence analysis of two-mode network data. Social Networks. 22(1):65-72 (2000).
- Roselló-Lletí E, Tarazón E, Barderas MG, Ortega A, Otero M, Molina-Navarro MM, Lago F, González-Juanatey JR, Salvador A, Portolés M, Rivera M. Heart mitochondrial proteome study elucidates changes in cardiac energy metabolism and antioxidant PRDX3 in human dilated cardiomyopathy. PloS one. 9(11):e112971 (2014).
- Saguner AM, Buchmann B, Wyler D, Manka R, Gotschy A, Medeiros-Domingo A, Brunckhorst C, Duru F, Mayer KA. Arrhythmogenic Left Ventricular Cardiomyopathy Suspected by Cardiac Magnetic Resonance Imaging, Confirmed by Identification of a Novel Plakophilin-2 Variant. Circulation. 132(6): e38-e40 (2015).
- Suárez E, Sariol CA, Burguete A, McLachlan G. A tutorial in genetic epidemiology and some considerations in statistical modeling. Puerto Rico health sciences journal. 26(4) (2007).
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic acids research. 43(D1):D447-D452 (2015).
- Tardiff, JC. Thin filament mutations developing an integrative approach to a complex disorder. Circulation Research. 108(6):765-782 (2011).
- Towbin JA, Lorts A, Jeffries JL. Left ventricular non-compaction cardiomyopathy. The Lancet. 386(9995): 813-825 (2015).

- Trinick J, Knight P, Whiting A. Purification and properties of native titin. *Journal of molecular biology*. 180(2): 331-356 (1984).
- Ullal AJ, Abdelfattah RS, Ashley EA, Froelicher VF. Hypertrophic Cardiomyopathy as a Cause of Sudden Cardiac Death in the Young: A Meta-Analysis. *The American journal of medicine*. 129(5):486-496 (2016).
- van Tintelen JP, Entius MM, Bhuiyan ZA, Jongbloed R, Wiesfeld AC, Wilde AA, van der Smagt J, Boven LG, Mannens MM, van Langen IM, Hofstra RM. Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/ cardiomyopathy. *Circulation*. 113(13): 1650-1658 (2006).
- van Rijsingen IA, van der Zwaag PA, Groeneweg JA, Nannenberg EA, Jongbloed JD, Zwinderman AH, Pinto YM, Dit Deprez RH, Post JG, Tan HL, de Boer RA. Outcome in phospholamban R14del carriers: results of a large multicentre cohort study. *Circulation: Cardiovascular Genetics*. CIRCGENETICS-113 (2014).
- Van Tintelen JP, Pieper PG, Van Spaendonck-Zwarts KY, Van Den Berg MP. Pregnancy, cardiomyopathies and genetics. *Cardiovascular research*. cvu014 (2014).
- Verducci JS, Melfi VF, Lin S, Wang Z, Roy S, Sen CK. Microarray analysis of gene expression: considerations in data mining and statistical treatment. *Physiological genomics*. 25(3):355-363 (2006).
- Walsh R, Rutland C, Thomas R, Loughna S. Cardiomyopathy: a systematic review of disease-causing mutations in myosin heavy chain 7 and their phenotypic manifestations. *Cardiology*. 115(1):49-60 (2010).
- Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, Mazzarotto F, Blair E, Seller A, Taylor JC, Minikel EV. Exome Aggregation Consortium, MacArthur DG, Farrall M, Cook SA, Watkins H. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *bioRxiv*: 041111(2016).
- Wasserman S, Faust K. *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press (1994).
- Yang J, Xu WW, Hu SJ. Heart failure: advanced development in genetics and epigenetics. *BioMed research international*. 2015 (2015).
- Yoshikawa T, Baba A, Nagatomo Y. Autoimmune mechanisms underlying dilated cardiomyopathy. *Circulation Journal*. 73(4):602-607 (2009).
- Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Human molecular genetics*. ddu125 (2014).

ABSTRACT

Research on Cardiomyopathy marker gene through gene network analysis

Myoung Seock Seo

**Laboratory of Computational Biology and Bioinformatics
Graduate School of Public Health
Seoul National University**

Cardiomyopathy is a disease that causes structural and functional abnormalities of the heart muscle without coronary artery disease, and the abnormalities of the cardiac chambers are abnormally uneven or enlarged. Cardiomyopathy can occur in all ages, and symptoms can vary from heart failure, arrhythmia, thrombosis, and sudden cardiac death. In addition, because the treatment varies greatly depending on the subtype and symptom, it is important to diagnose it early to ensure effective treatment and management. Recent advances in sequence analysis and bioinformatics techniques have led to an increase in the amount of genetic information associated with cardiomyopathy, as well as the ability to analyze a much larger number of genes in a short period of time than conventional sequencing methods, it helps diagnose cardiomyopathy by providing an overall genetic analysis of the disease. In order to utilize genetic information efficiently, it is important to construct a database and a search system. Therefore, in this study, I constructed a genetic database and a web-based search system for cardiomyopathy and named it CGDB (Cardiomyopathy Genetic Database).

GenBank, ClinVar, and UniProt were used to obtain information on gene and amino acid sequence, gene mutation, clinical significance, and protein used in database construction. The downloaded flat file was subjected to a data parsing process using JAVA programming language, and re-processed it into a file in sql format to input it into Mysql. The database is implemented through a web - based search system using JSP and HTML. The implementation of this database can be divided into two major parts. First, it provides integrated genetic information for cardiomyopathy. It can search for genes satisfying the condition and obtain information about the mutation of the gene and the expressed protein. Secondly, it is possible to output sequence information of base and amino acid by fasta format file for the second study. Genetic information on cardiomyopathy provided through these databases and search systems may contribute to the rapid and accurate diagnosis of cardiomyopathy. Next, gene network analysis was performed to find marker genes for cardiomyopathy. First, I used NCBI GEO microarray data to select the genes that constitute the network. Among the patients with cardiomyopathy of various etiologies, the results of the patients showing the prognosis of heart failure were used to select differentially and commonly expressed genes to constitute a network of genes involved in cardiac failure as a prognosis of cardiomyopathy. The interaction between genes was constructed based on the expression correlation between genes. Information on expression correlations was obtained through the GeneNetwork web service. The results of centrality analysis about gene network suggest that C1QTNF7, ECM2, FAM188A, HNMT, ITIH5, NAMPT, OIP5-AS1, PTN, SMOC2 and SULF1 play an important role in the prognosis of heart failure. In addition, constructed 2-mode network between above network and the causative genes of congenital cardiomyopathy, and perform centrality analysis to select the causative genes of congenital cardiomyopathy which affect the expression of showing the prognosis of heart failure. As a result of centrality analysis, it was found that in dilated cardiomyopathy, ACTC1, ACTN2, BAG3, DES, DMD, DSG2, DSP, FKTN, ILK, LAMP2, LDB3, LMNA,

MYH7, MYPN, PKD1, PKP2, PLN, PSEN1, RBM20, RYR2, SCN5A, SGCD, TCAP, TPM1, TTN and VCL, and in hypertrophic cardiomyopathy, ACTC1, ACTN2, ANKRD1, LAMP2, MYOZ2, MYPN, PLN, PRKAG2, TTN and VCL were found to play an important role in the expression of genes showing the prognosis of heart failure. It can be used as a marker gene to predict the prognosis of heart failure in a diagnostic test using peripheral blood. The results of this study suggest that the accurate diagnosis and prognosis of heart failure can be predicted, which can contribute to the prolongation of life and the quality of life of patients through appropriate treatment direction and continuous monitoring.

.....
keywords : cardiomyopathy, bioinformatics, database, network analysis,
marker gene, public health

Student Number : 2015-24008